EB

ENCICLOPEDIA DE BUZUNAR

A. A.
PERVOZVANSKI

Căutarea deciziei optime

EDITURA ENCICLOPEDICĂ ROMÂNĂ

În memoria lui
Alexandru Aronovici Feldbaum

Cuvînt înainte

Cărțile de popularizare a științei sînt de diferite feluri. Unele din ele nu reprezintă decît o extindere a rubricii „Știați că ...?" *pe care o găsim în anumite reviste; în ele munți de fapte se înalță unii peste alții, legăturile dintre acestea reducîndu-se la comparații de felul următor:* „**A** *este mai mare decît* B *de un milion de ori, iar* B *este, la rîndul său, mai mare decît* C *tot de un milion de ori; dacă* **A** *ar fi încărcat în vagoane, trenul respectiv s-ar întinde pînă la steaua Proxima...*" *și altele asemenea. Din fericire, literatura de acest gen începe să stîrnească antipatie în mase tot mai largi și, treptat, părăsește arena. În ultima vreme au început să fie la modă cărți de un alt gen, în care știința este de fapt canavaua, pretextul unor raționamente generale, a unor eseuri istorico-filozofice. Asemenea cărți sînt scrise de către savanți autentici care și-au descoperit talent literar sau de către autentici oameni de litere care și-au dat seama de farmecul științei și al savanților. Cărțile bune de acest gen produc o adevărată desfătare. Lectura lor este tot atît de plăcută cum e conversația cu un om spiritual și erudit care dispune totdeauna de un citat interesant sau de o anecdotă plăcută. E drept, ciocnindu-se, vrînd-*

nevrînd, de necesitatea de a se referi în discurs la știința propriu-zisă, în întreaga ei nuditate, calmă și aspră, autorii unor asemenea cărți încearcă o anumită jenă: merită oare să obosească atenția cititorului cu amănunte și subtilități, oare nu e mai bine să expedieze toate acestea într-un apendice, care tot nu va fi citit, sau pur și simplu să le înlăture? Asemenea contradicții, independent de modul în care sînt rezolvate de fiecare autor în parte, conduc la aceeași consecință: indiferent dacă în textul de bază vor rămîne sau nu vor rămîne amănuntele, în conștiința cititorului ele vor pătrunde ca niște lucruri foarte neplăcute, a căror prezență pare să se datoreze unei neînțelegeri. Generalizările filozofice ale autorilor sînt atît de îmbelșugate și promițătoare, ultima anecdotă povestită este atît de scînteietoare, încît pur și simplu n-ai cînd să te mai oprești la o teoremă sau la o figură, inima te îndeamnă să citești cît mai repede în continuare, doar te așteaptă lucruri atît de atractive! Este posibil, desigur, să existe și cititori deosebit de migăloși, care să acorde interes și amănuntelor, dar, cinstit vorbind, ei îmi inspiră milă, căci pierd cea mai mare parte din plăcerea pe care o oferă asemenea cărți interesante și distractive.

Și iată că a venit momentul potrivit să recunosc că această carte, pe care ați deschis-o cu cîteva minute în urmă, deși își zice „de popularizare", este departe de a fi atît de scînteietoare și atrăgătoare. Ea este **o lucrare de informare de tipul „introducerilor"; ea vă introduce într-o disciplină destul de complicată: teoria căutării deciziilor optime, teoria alegerii din numeroase acțiuni posibile a aceleia care va aduce utilitatea maximă.** *Dacă cunoașteți cît de cît conglomeratul de științe care în mod convențional este denumit* **cibernetică,** *atunci probabil că știți că* **problema optimizării** *este una dintre problemele cele mai importante în știință în general și în cibernetică în special. Dacă însă nu știți nimic despre toate acestea, dar sînteți dotați cu o curiozitate naturală (care, după cum se va lămuri din cele ulterioare, este condiția fundamentală pentru o bună întîlnire cu cartea), atunci parcurgeți capitolul 1, care are de fapt rolul de introducere în materie.*

În cele ce urmează aș vrea să spun cîteva cuvinte celor cărora tema căutării deciziei optime li se va părea interesantă. **Această carte are un dublu scop: cultural și utilitar. Ea poate servi și drept manual,** *întrucît autorul a urmărit nu numai trezirea interesului pentru anumite probleme științifice, ci a vrut să dea cititorului și posibilitatea de a gusta din roadele practice ale științei, învățîndu-l cîteva metode și procedee fundamentale. Acestui scop îi servește și modul în care este construită lucrarea, deoarece ea merge de la mai simplu la mai complicat, descoperind treptat nivele noi de înțelegere a problematicii* . Firește, ea nu este totuși un manual cu adevărat „serios", întrucît este destinată unui cerc cît mai larg posibil de cititori, cu un nivel inițial de pregătire variat, pornind, de exemplu, de la nivelul celor care nu au terminat decît primele opt clase ale școlii generale. Iar țelul pe care și-l propune — de a nu se limita la indicarea „strălucitoarelor culmi" ale științei moderne, ci de a ajuta pe cititor să ajungă la ele pe cărări logice adeseori abrupte, impune dotarea expunerii principale cu diferite suporturi, care iau aspectul unor prezentări fugitive și întrucîtva superficiale ale unor capitole ale științei, mai ales ale matematicii, care depășesc cunoștințele cititorului cu nivelul de pregătire amintit. Adevărul este că, chiar dintre specialiști, mulți și-au încheiat studiile superioare într-o vreme cînd matematica era nesatisfăcător predată și încă mai nesatisfăcător receptată, deoarece părea inutilă pentru un inginer practician și cu atît mai mult pentru un biolog sau un economist. Apariția mașinilor de calcul a produs o revoluție în opiniile cu privire la utilitatea practică a matematicii. Numeroși foști sceptici și-au dat seama în mod clar că modelarea matematică a fenomenelor reale este posibilă, dar ei nu mai sînt în stare acum să depună un efort care cere un mare volum de muncă pentru a-și asimila conținutul unor manuale voluminoase, ai căror autori, de regulă, nu se prea îngrijesc de accesibi-*

---

* În carte se pot demarca două părți: prima parte (capitolele 1—11) conține materialul de bază; a doua (capitolele 12—15) — un material auxiliar, care se asimilează întrucîtva mai dificil.

litatea materialului pentru cititorii cu „defecțiuni" în pregătirea teoretică. Desigur că lectura acestei cărți nu poate înlocui integral studiul sistematic al problemelor pe care ea le abordează; ea poate și trebuie totuși să joace atît rolul de prim teritoriu cucerit, cît și de cap de pod pentru cuceriri ulterioare. Tocmai aceasta a stimulat eforturile de a construi cartea nu ca un ghid-compendiu, un ansamblu de rețete gata de a fi utilizate, dar cu rezultate despre care nu se știe de unde provin, ci ca o construcție logică, care dezvăluie „laboratorul de idei" în care se făuresc noile concepții științifice și, mai ales, simplitatea lăuntrică a acestora. Aș vrea să sper că în virtutea acestui fapt răsfoirea cărții nu va fi lipsită de interes chiar și pentru specialiștii care se ocupă nemijlocit cu **teoria căutării** *și aplicațiile sale.*

Cartea a fost sensibil îmbunătățită în urma sfaturilor primite din partea prof. I. B. Celpanov și din partea altor colegi ai autorului de la Institutul politehnic din Leningrad. Foarte utilă a fost și critica amănunțită a manuscrisului efectuată de către prof. A. I. Lerner, critică de care am ținut seama la redactarea variantei finale. Un mare ajutor în executarea calculelor și pregătirea materialului grafic l-am primit de la M. B. Ulițki. Autorul este sincer îndatorat tuturor acestor persoane.

Cartea a fost concepută și scrisă într-o vreme cînd marele specialist sovietic în teoria și aplicațiile căutării deciziei optime, Alexandru Aronovici Feldbaum, era deja grav bolnav, astfel încît autorul nu a putut beneficia de sfaturile sale. Dar tocmai cunoașterea lucrărilor lui A. A. Feldbaum, iar mai tîrziu întîlnirea cu el personal au trezit la autor, ca și la mulți alți tineri cercetători, interesul pentru acest cerc de probleme. Dedicînd cartea luminoasei sale amintiri, autorul plătește doar un mic tribut acelei profunde stime de care s-a bucurat în rîndurile noastre acest remarcabil om și savant.

1
Intuiție și logică

(Schema fundamentală a căutării)

Ceva uman rămîne-n orice lucru
Ce-a fost atins de mînă omenească.
S. Marșak

Căutarea este unul din fenomenele cele mai caracteristice ale activității umane. Ne-am obișnuit cu expresii atît de radical diferite ca sens, cum ar fi „căutarea minereurilor", „căutarea erorilor", „căutarea unui rost în viață", „căutarea construcției celei mai bune", și nu ni se pare deloc ciudată prezența în fiecare din ele a unuia și aceluiași cuvînt. Evident, intuitiv ne sînt clare acele trăsături comune care ne permit să unificăm diferite situații din viața de toate zilele, din știință și din tehnică în conceptul de *căutare*. Merită însă să le supunem unei analize logice și să degajăm cu claritate esențialul.

În primul rînd, este clar că în toate situațiile amintite se caută, se urmărește ceva. *Prin urmare, căutarea presupune prezența unui* **scop**. Tocmai denumirea scopului constituie atributul determinant al cuvîntului „căutare" în exemplele date.

În plus, este necesar să observăm că nu se poate vorbi despre căutare dacă scopul propus este imediat accesibil. Căutarea presupune o mișcare în direcția țelului, ea este un *proces* care se desfășoară în timp și reprezintă o succesiune de acțiuni.

Mai mult, înaintarea către țel nu este căutare dacă drumul este unul singur, cunoscut dinainte. *Prezența unei mulțimi de drumuri posibile, necesitatea* **alegerii** *unuia dintre ele, iată o trăsătură esențială a oricărui proces de căutare.*

În sfîrșit, *orice căutare rațională se realizează în concordanță cu un anumit* **plan** *care reprezintă o* **regulă de alegere** *a acțiunilor succesive; acesta este* **algoritmul** *căutării. Tocmai* **modalitatea de construire a acestei reguli** *este sufletul căutării, reflectînd* **natura principială** *a procesului, indiferent de scopurile concrete, care pot merge de la prinderea unui criminal pînă la determinarea celui mai bun regim de funcționare a unei mașini.*

Iată de ce conținutul principal al cărții este consacrat descrierii diferiților algoritmi ai căutării în acea reprezentare sintetică pe care o permite limbajul matematicii și al ciberneticii.

Dar înainte de a pune în funcțiune această puternică armă, să încercăm să urmărim pe o cale pur logică unele idei ușor de înțeles, dar esențiale pentru organizarea căutării.

Este evident că alegerea oricărei variante de acțiune este încununată de un anumit *rezultat,* care reprezintă încheierea acțiunii respective.

În unele cazuri acest rezultat este doar un purtător de informație elementară: scopul este atins sau nu este atins. Dacă este atins, atunci firește, nu mai avem ce să căutăm mai departe, procesul este încheiat; iar dacă nu este atins, la prima vedere s-ar părea că acțiunea întreprinsă a fost inutilă. În realitate însă, lucrurile nu stau așa; rezultatul a fost că una dintre variante, fiind testată, a eșuat și deci numărul variantelor posibile s-a redus; așadar, problema căutării ulterioare a devenit mai simplă.

Să dăm un exemplu tehnic deosebit de important, în cadrul căruia se face uz tocmai de o asemenea informație elementară.

Mașinile de calcul cifric (calculatoarele) reprezintă una din cele mai mari realizări ale tehnicii moderne,

„minunea secolului al XX-lea". Ele sînt constituite din mii de piese distincte: rezistențe, condensatoare, elemente semiconductoare, îmbinate în scheme foarte complicate. Numărului mare de piese și complexității conexiunilor lor li se datorește în mare măsură faptul că aceste mașini pot fi pe drept cuvînt denumite „inteligente", dar, pe de altă parte, tot aici rezidă cauza neplăcerilor pe care le întîmpină inginerii ce lucrează cu calculatoarele. Într-adevăr, ieșirea din funcțiune chiar și a unei singure piese poate să aibă ca rezultat faptul că mașina „inteligentă" începe să dea rezultate cu totul absurde. Căutarea defecțiunii, a cauzei pentru care calculul „eșuează" este o întristătoare necesitate. Cum să-l identifici în mulțimea de blocuri și piese pe vinovatul principal? Se face apel la un program de testare: mașina este pusă să rezolve o problemă la care răspunsul corect este cunoscut și despre care se știe ce anume blocuri ale mașinii sînt solicitate în procesul rezolvării. Dacă programul test a decurs cu bine, atunci, firește, defecțiunea nu a fost descoperită; în schimb au fost verificate dintr-o dată numeroase blocuri care nu mai sînt suspecte și deci se poate trece la verificarea altora. Dacă însă răspunsul obținut nu este cel corect, defectul trebuie căutat undeva în interiorul grupei de blocuri respective. În acest caz se poate face apel la un nou test, care cuprinde numai o parte din ceea ce este suspect, reducînd astfel și mai mult numărul localizărilor posibile ale defecțiunii. Dacă sînt organizate rațional, aceste operații repetate conduc infailibil la un succes deplin.

Exemplul considerat este foarte caracteristic, însă de regulă **studiul unei variante dă o informație destul de bogată, permițînd nu numai reducerea numărului posibilităților, dar și indicarea direcției în care trebuie continuată căutarea, indicarea variantei care urmează să fie testată.**

Astfel, căutarea amintește adesea jocul de-a „cald și rece": cu cît obiectul ei este mai aproape, cu atît este mai „cald", cu cît ne depărtăm de țel, cu atît este mai „rece"; iar rolul de „călăuză" îi revine chiar rezultatului

variantei alese. Un exemplu simplu: dacă în urma unei mici creșteri a alimentării cu oxidant randamentul unui motor a devenit ceva mai mare — lucru echivalent cu semnalul „mai cald" —, asta înseamnă că putem să mai intensificăm întrucîtva alimentarea cu oxidant și poate că vom avea „și mai cald"; dar poate că, dimpotrivă, instrumentele de măsură vor indica o reducere a randamentului. Înseamnă că în procesul căutării „am sărit" peste acea variantă pentru care coeficientul de eficiență este maxim; este indicat așadar să ne întoarcem puțin înapoi.

Aici ne lovim de o particularitate foarte importantă a majorității proceselor de căutare care se înfăptuiesc în tehnică (și nu numai în tehnică). De obicei, aceste procese se întemeiază pe principiul: „operații apropiate dau efecte apropiate". Dacă în locul unei variante adoptăm alta care diferă puțin de ea, atunci ori ne vom apropia puțin de țel, ori ne vom depărta de el, dar tot puțin. Să presupunem că ne-am apropiat de țel. Atunci confruntarea celor două variante considerate, adică explicitarea deosebirilor dintre ele, ne permite să lămurim tendința, direcția mișcării către țel. Se poate presupune că adoptarea unei variante care diferă și mai mult de cea inițială în privința acelorași particularități ne va apropia și mai mult de țel.

Așadar, lămurirea consecințelor ce corespund unei variante sau unor variante pentru care am optat ne permite să decidem care variantă nouă poate fi încercată pentru a ne apropia de țel.

Rezultatul încercării nu este cunoscut dinainte. El trebuie aflat fie prin calcul, dacă problema este formulată matematic, fie prin organizarea unui experiment cu observațiile și măsurătorile corespunzătoare. Ca în povestea cu prostuțul șoricel căruia i se căuta cea mai bună doică: „Mamă, nu mai chiți, mai bine doică mi-ai găsi! Mama iute a alergat și doică pe rață a luat"...

Așadar, rața este prima variantă. Se testează această variantă, are loc primul experiment. Dar rezultatul este nesatisfăcător: „Nu, rață, nu-mi ești pe plac, prea

e tare al tău mac-mac". Varianta a doua este broasca, dar: „Nu, broască, nu-mi eşti pe plac cu al tău orac-uac" şi aşa mai departe.

Avem de-a face cu o schemă tipică a produsului căutării: variantă, experiment, selecţie conform cu rezultatul. Ce-i drept, şoricelul n-avea un criteriu clar de selectare şi nici un plan clar de acţiune. După cum este bine cunoscut, căutarea s-a încheiat pentru el în modul cel mai trist: i-a plăcut mieunatul pisicii. Dar aceeaşi lipsă de claritate, în pofida încercărilor de a ieşi din tufişul intuiţiei, este din păcate caracteristică şi acţiunii descrise de noi. Am vorbit mai sus despre asemenea lucruri ca „efect", „apropiere de ţel", „variante apropiate", dar ce sînt toate acestea n-am definit exact. Trista soartă a şoricelului care s-a încrezut în intuiţia spontană ne cheamă încă o dată la precizie şi la claritatea noţiunilor. Aşadar...

2
Către ce tindem?

(Noțiunea de funcție-scop)

Și binele și răul vin pe neașteptate,
Iar celui ce le prevestește
nu-i dă crezare nimeni.
După **Goethe** *(Faust)*

Problema pe care o pune titlul are o formulare atît de profundă în aparență și atît de vagă de fapt, încît s-ar părea că și răspunsul trebuie să fie deosebit de vag.

Să restrîngem însă sensul acestei întrebări: nu vom căuta răspunsul la întrebarea care sînt țelurile posibile ale omului sau ale omenirii, ci vom descrie doar acea formă — unica posibilă după părerea noastră — în care poate fi exprimat orice scop, desigur dacă este conceput în mod clar (în ce privește neclaritatea, reveniți la povestea prostuțului șoricel).

Îmi voi permite, așadar, să formulez ipoteza de bază: scopul oricărei căutări constă în a găsi ceva mai bun, în a trece de la o stare la alta mai bună, preferabilă pentru subiectul care întreprinde căutarea.

Orice căutare presupune însă existența unei mulțimi de variante, căi sau stări. Axioma adoptată afirmă că testînd două variante oarecare se poate aprecia, fie și subiectiv, care dintre ele dă un rezultat mai bun, sau cel puțin se poate afirma că ambele variante sînt echivalente.

Presupunem așadar că mulțimea variantelor este înzestrată cu o anumită ordine interioară, unele dintre variante

fiind mai bune, altele mai rele, deși această ordine nu este cunoscută dinainte (adică înainte de executarea experimentelor corespunzătoare care constau în testarea variantelor). Vom aminti că dacă o asemenea ordine ar fi cunoscută, necesitatea căutării ar dispărea: s-ar putea trece dintr-o dată la adoptarea variantei optime.

Astfel, la prima vedere, am ajuns la o concepție destul de substanțială: **scopul căutării, adică ceea ce constituie obiectivul către care tindem, rezidă în selectarea celei mai bune dintre numeroase variante comparabile prin rezultatele lor.** Dar „ce e rău și ce e bine?" La această întrebare răspundem cu hotărîre: mai bine este ceea ce este mai mult. „Dar de ce?" va obiecta cititorul indignat. „Ar fi oare mai bine să existe mai multe războaie, mai multe boli, mai multe suferințe?!" Ca să nu ne acuzați de militarism și antiumanism, răspundem: „Fie, sîntem de acord. Mai bine înseamnă mai puțin." Dar riposta este o indignare și mai mare: „oare să fie mai bine mai puțină hrană, mai puțină îmbrăcăminte și așa mai departe (conform cunoscutei liste a bucuriilor pămîntești)?!" Astfel, autorul se vede nevoit să urmeze marele exemplu al lui Nastratin Hogea: „Și tu ai dreptate, dar și tu ai dreptate". Uneori mai bine înseamnă mai mult, alteori mai bine înseamnă mai puțin, dar în fiecare proces concret de căutare ne putem înțelege ce este mai bine: mai puțin sau mai mult. Esențial este că ne-am înțeles cu privire la necesitatea de a stabili o corespondență între rezultate și numere, necesitatea de a atașa un anumit număr rezultatului fiecărei variante în parte. Comparînd două numere, vom putea atunci spune care dintre cele două variante este mai bună. Regula de selecție a variantelor se va putea deci exprima riguros matematic. Să zicem, de exemplu, că ne-am înțeles: mai bine înseamnă mai mult. Atunci, dacă verificînd varianta 1 obținem rezultatul exprimat prin numărul a, iar încercînd varianta 2 obținem numărul b, putem spune: dacă a este mai mare decît b, atunci este mai bună varianta 1, dacă a este mai mic decît b, atunci este mai bună varianta 2, iar dacă a este egal cu b, ambele variante sînt echivalente.

Aceasta reprezintă deja un rezultat serios. Forța lui rezidă mai ales în aceea că se deschide perspectiva de a încredința calculatoarelor selectarea variantelor; desigur, calculatoarele nu știu ce e rău și ce e bine, dar se pricep minunat (și mult mai repede decît omul) să răspundă la întrebarea ce e mai mult și ce e mai puțin. Avînd în vedere uriașa cantitate de variante posibile în procesele reale de căutare, rapiditatea răspunsului este destul de importantă, și cititorul se va convinge de aceasta în repetate rînduri.

Pe de altă parte, rezultatul obținut modifică însuși caracterul expunerii: din acest moment critic vom putea trece de pe calea nesigură a considerentelor etice și filozofice pe drumul lipsit de surprize al metodelor matematice. Mai înainte de toate să amintim că unul dintre conceptele matematice fundamentale este conceptul de *funcție*. Să presupunem că avem o mulțime (colecție) arbitrară de elemente și fiecărui element îi este atașat un anumit număr. Vom spune atunci că ne este dată *o funcție definită pe mulțime*, iar mulțimea însăși se numește *domeniu de definiție* (sau *sursă*). Astfel, atașînd fiecăreia dintre variantele posibile un anumit număr, noi am definit prin aceasta o anumită funcție pe mulțimea variantelor, o funcție care de obicei se numește *funcție-scop**. Domeniul de definiție al funcției-scop este denumit adesea *domeniul căutării*. Atunci cînd pornim căutarea, de obicei nu cunoaștem forma explicită a funcției-scop, dar putem determina diferitele ei valori, efectuînd verificarea variantei alese. În procesul căutării tindem în mod uzual să aflăm varianta cea mai bună, adică acea variantă căreia îi corespunde valoarea cea mai mare (sau cea mai mică) a funcției-scop, deși adesea ne declarăm satisfăcuți și cu asemenea variante care sînt doar mai bune decît cea inițială. În orice caz, țelul „sacru" este obținerea variantei celei mai bune, optime; tocmai de aceea problemele căutării sînt denumite adesea *pro-*

* În literatura de specialitate se mai utilizează denumirea de *funcție-obiectiv*, mai rar *funcție-criteriu*, *funcție de optimizare* (*N.T.*).

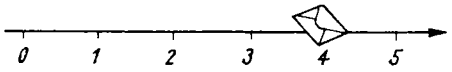


Fig. 2.1

bleme de optimizare. În matematică, cea mai mare valoare a unei funcții se numește *maxim*, cea mai mică *minim*, ambele fiind încadrate în denumirea comună de „extremum"; de aci provine o altă denumire a problemelor căutării: *probleme de extremum* sau *extremale*; tot astfel se explică titlurile multor tratate științifice: *Despre metodele de căutare a extremumului, Noi metode de căutare a extremumului, Alte metode de căutare a extremumului* etc.

Așadar am introdus o ordine matematică în rezultatele încercării diferitelor variante. Dar ce să facem cu variantele înseși, cum să introducem ordine în această mulțime de posibilități?

Rețeta generală este aceeași ca și mai înainte: fiecărei variante trebuie să-i atașăm un anumit număr sau cîteva numere*. Să dăm un exemplu simplu. Dacă un uituc are ceva spirit metodic, atunci, pornind să caute un plic rătăcit undeva în sertarele unei mese, el începe să urmărească sertarele unul după altul, de exemplu începînd cu cel mai de jos din stînga, după care trece la al doilea de jos și așa mai departe; astfel, el ordonează mulțimea variantelor. Ar fi așadar firesc ca sertarului cel mai jos din stînga să-i dăm numărul de ordine 1, următorului numărul 2 etc. și să reprezentăm mulțimea variantelor (sertarelor) sub forma unor puncte pe axa numerică (fig. 2.1).

Să presupunem că plicul căutat se află în sertarul al patrulea (ceea ce uitucul firește că nu știe dinainte). Întrucît, cu excepția variantei 4, toate variantele sînt echivalente (plicul nu se află în nici unul din sertare), numerele (valorile) atribuite lor trebuie să fie identice.

---

* Uneori e necesar să atașăm chiar o funcție, dar cu asemenea probleme (de căutare în spații funcționale) nu ne vom ocupa aici, din cauza complexității lor.

De exemplu, să presupunem că ele sînt toate egale cu unitatea (aici nu contează mărimea). Dar variantei 4 trebuie să-i atribuim alt număr. Dacă admitem că „mai bine înseamnă mai mult", adică uitucul caută un maxim, atunci variantei 4 i se poate atribui, de exemplu, numărul 2. Funcția-scop ia forma indicată în figura 2.2, unde maximul corespunde variantei 4.

În exemplul nostru, funcția-scop este definită numai într-un număr finit de puncte (întrucît, din fericire, nu există mese cu un număr infinit de sertare); o asemenea funcție este de obicei denumită discretă. Modul de construire a funcției este destul de artificial și arbitrar, dar un asemenea arbitrar în formularea matematică a problemelor căutării este pe deplin uzual. Poate va fi mai accesibil un alt exemplu, în cadrul căruia funcția-scop ne apare într-un mod mult mai firesc. Ai învîrtit butonul aparatului de radio căutînd un post interesant; ai și găsit acest post, dar el abia dacă se aude, deși amplificatorul este pus la extrem. Atunci este necesar să precizezi mai bine direcția aparatului: rotindu-l ceva mai la stînga, se aude mai bine, dacă-l rotești încă puțin se aude și mai tare, dar, dorind și mai binele, l-ai mai rotit puțin la stînga și iată că sunetul a devenit mai slab. Atunci trebuie să revii, încercînd poziția cea mai bună a aparatului, căreia îi corespunde sunetul cel mai puternic. Este un proces tipic de căutare. Mulțimea variantelor este aici mulțimea pozițiilor aparatului (a unghiurilor φ), ea este ordonată în mod firesc de la pozi-

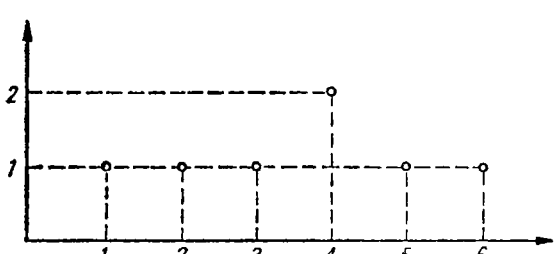


Fig. 2.2

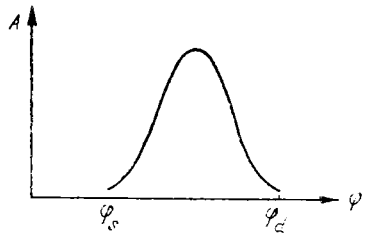


Fig. 2.3

ția de extremă dreaptă φ_d, în care postul abia se mai aude, pînă la poziția extremă stîngă φ_s, unde sunetul este iarăși slab. Funcția-scop este dependența sunetului A de unghiul φ. O asemenea funcție poate să arate, de exemplu, ca în figura 2.3.

În acest exemplu se poate înțelege mult mai ușor ce reprezintă funcția-scop, dar căutarea însăși a maximului dorit este mai complicată. Complicarea constă în aceea că aici numărul variantelor este infinit: orice poziție a aparatului între unghiurile extreme φ_d și φ_s poate fi bănuită de a fi extremală; enumerarea tuturor acestor poziții sau, ceea ce revine la același lucru, a tuturor punctelor de pe segmentul de la φ_d pînă la φ_s este imposibilă. De aceea, nu este posibilă nici simpla triere, care corespunde examinării succesive a variantelor (sertarelor) din exemplul precedent. Pentru determinarea exactă a extremumului este necesar să facem uz de alte procedee; unul din ele îl și realizăm intuitiv rotind aparatul. Totuși, cele două exemple descrise aici se aseamănă prin aceea că în ambele probleme mulțimea variantelor posibile poate fi reprezentată sub forma unei mulțimi de puncte de pe o dreaptă. Într-un anumit sens, asemenea probleme sînt simple; ele sînt numite *unidimensionale* (dreapta este un spațiu unidimensional, planul este bidimensional, corpurile sînt părți ale spațiului tridimensional etc.).

Dar în practică avem de-a face cel mai adesea cu probleme de o natură mai complicată. Ca exemplu, să examinăm problema optimizării producției de cauciuc sin-

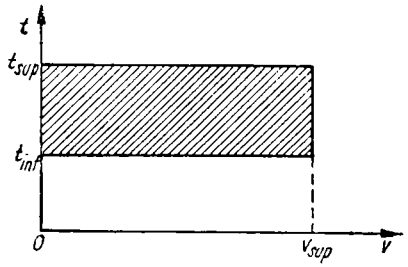


Fig. 2.4

tetic, un polimer complex, din hidrocarburi simple, monomeri (divinilul și stirenul), proces care are loc în baterii polimerizatoare*. Cursul reacțiilor chimice în aceste baterii depinde de mulți factori: de viteza de alimentare cu materia primă de bază și de compoziția acesteia în procente, de alimentarea cu chimicalele auxiliare (oxidanți, reducători etc., în total peste zece), de temperatura procesului și de viteza în care are loc amestecul. Eficiența procesului este determinată de obicei de gradul de conversiune, adică de procentul de molecule monomere care s-au unit în molecule polimere. Țelul căutării regimului optim este obținerea gradului maxim de conversiune prin alegerea unei îmbinări corespunzătoare a factorilor de care depinde mersul reacției.

Să presupunem că factorii pe care îi putem varia după voie sînt numai în număr de doi, de exemplu temperatura și viteza procesului de amestecare, ceilalți factori fiind dați dinainte. Atunci un regim posibil va fi determinat de două numere, t și v (valorile temperaturii și vitezei), ceea ce se poate reprezenta geometric printr-un punct din plan cu coordonatele t, v (fig. 2.4); în acest caz, mulțimea regimurilor posibile, adică ceea ce am denumit domeniul căutării, se reprezintă prin dreptunghiul hașurat mărginit de dreptele corespunzătoare valorilor limită ale

* Autorul a extras această problemă din lucrarea lui N.I. Gherasimov.

temperaturii (t_{inf} și t_{sup}) și valorilor limită ale vitezei procesului de amestecare (0 și v_{sup}), admisibile prin condițiile tehnologice. Căutarea regimului optim se va reduce așadar la căutarea acelui punct din dreptunghiul considerat, adică a acelei asocieri pentru t și v, pentru care gradul de conversiune va fi cel mai mare.

O asemenea problemă de căutare, care depinde de doi parametri, se numește *bidimensională*. Dacă ar exista posibilitatea alegerii și pentru alți factori care influențează mersul procesului, de exemplu componența în procente a amestecului inițial, problema căutării ar fi tridimensională, cvadridimensională etc.

Urmînd tradițiile strămoșilor noștri, oamenii primitivi, vom întrebuința cuvîntul „mult" pentru tot ce este mai mare decît unitatea și deci toate problemele care nu sînt unidimensionale le vom denumi *multidimensionale*.

Dar cum să facem față acestei nenumărate mulțimi de variante, cum să construim un sistem rezonabil de căutare? După cum am mai spus, este posibil să le enumerăm pe rînd. Va trebui să reflectăm la alte proceduri, dar despre aceasta vom mai vorbi.

Deocamdată, să tragem cîteva concluzii.

Ne-am înțeles că problemelor căutării li se poate da o formă matematică riguroasă dacă atașăm fiecăreia dintre variantele posibile un număr sau o colecție de mai multe numere (în cazul precedent au fost două numere, temperatura și viteza) și dacă ținem seama că și rezultatul încercării fiecărei variante în parte (adică rezultatul experimentului respectiv) se poate asocia cu un anumit număr, care înainte de executarea experimentului este necunoscut. Am aflat că în felul acesta se determină funcția-scop pe domeniul căutării, după care problema căutării poate fi considerată ca problema aflării extremumului acestei funcții.

În încheiere trebuie să răspundem unei întrebări „afurisite" care va apărea inevitabil în mintea unui cititor inteligent și cîrcotaș: „Oare întotdeauna se poate da o formă numerică problemelor reale de căutare? Doar nu se poate descrie printr-un număr un sentiment! " La această întrebare trebuie să răspundem deschis: „Din

păcate, sau din fericire, este adevărat că nu întotdeauna". În multe cazuri sîntem în situația copilului care privește pierdut pe nenea ce-i pune întrebarea: „Pe cine iubești tu mai mult, pe tata sau pe mama?" Nu, nici copilul, nici noi, care sîntem pe deplin adulți, nu vom putea răspunde la asemenea întrebări elementare, nu vom putea organiza comparații, ordonări și căutări ale ființei celei mai iubite, ale cîntecului preferat, ale cărții care ne captivează mai mult decît oricare alta, ale locului favorit pentru plimbări și așa mai departe.

Sentimentele sînt realmente neformalizabile, dar cui îi trebuie dragostea formalizată? Se poate, ce-i drept, constitui o schemă de căutare a „tovarășului de viață" optim, dar aici nu va fi loc pentru sentimente. Dar cu excepția acestui domeniu poetic și sacru, viața noastră este plină de nenumărate probleme de căutare care pot fi formalizate cu precizie fără a pierde prin aceasta nimic din splendoarea lor.

3
Peşte mare, peşte mic, prinde-te în plasă

(Căutarea pe reţea prin triere)

Experienţa — iată temelia cunoaşterii.
Alcmanos (secolul al VII-lea î.e.n.)

După ce ne-am lovit de probleme în care mulţimea variantelor nu poate fi numărată, am fost nevoiţi să renunţăm la ideea atotputerniciei trierii directe. Trebuie să născocim alte căi. Acestea sînt sugerate de principiul deja cunoscut: operaţii apropiate conduc la efecte apropiate (vezi cap. 1). Înarmîndu-ne însă cu limbajul matematic, putem să formulăm acest principiu într-un mod ceva mai riguros: unor puncte apropiate în domeniul căutării le corespund valori apropiate ale funcţiei-scop. Cititorul care ştie ceva matematică va observa că acest principiu nu-i decît afirmaţia că funcţia-scop este continuă. Desigur, nu orice funcţie este continuă, dar, de regulă, funcţiile-scop care apar în mod natural în problemele reale de căutare au această proprietate.

Mai mult, de obicei aceste funcţii se comportă lin şi regulat, fără a face cotituri bruşte şi fără a prezenta oscilaţii rapide. Pe scurt, ele se comportă ca o domnişoară bine crescută care şi-a însuşit bunele maniere, şi nu ca un ştrengar despre care nu ştii niciodată dacă, după un minut de comportare liniştită, nu va face vreo boacănă.

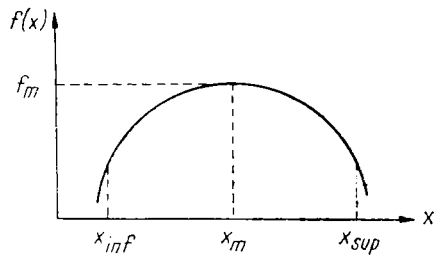


Fig. 3.1

Firește, în probleme apar adesea și funcții-strengar, care pricinuiesc cercetătorilor nu mai puțin sînge rău decît copiii neascultători mamelor lor. Dar despre aceasta vom mai vorbi. Deocamdată să examinăm funcțiile cu bună creștere, line și liniștite.

Să începem prin a examina problema căutării unidimensionale. În figura 3.1 este reprezentată o funcție-scop. Valoarea maximă a funcției, f_m, corespunde variantei căreia i s-a atribuit numărul x_m, iar domeniul căutării se întinde de la x_{inf} pînă la x_{sup}. Dacă acest tablou l-am avea în fața ochilor chiar de la început, căutarea ca atare n-ar mai fi necesară.

În realitate însă, de regulă avem posibilitatea de a măsura numai unele valori ale funcției, organizînd experiențe de verificare a diferitelor variante.

Să procedăm sistematic și să începem prin a elabora planul experiențelor. Vom divide întregul segment de la x_{inf} pînă la x_{sup} în mai multe părți egale, de exemplu zece (fig. 3.2). Punctele de diviziune le vom nota cu $x^{(1)}$; $x^{(2)}$; ... $x^{(9)}$. De asemenea, vom pune $x_{inf} = x^{(0)}$ și $x_{sup} = x^{(10)}$*. Aceste puncte se numesc *nodurile rețelei*, iar distanța dintre ele se numește *pasul rețelei* (în exemplul nostru este evident că pasul rețelei este $\frac{x_{sup} - x_{inf}}{10}$).

Să presupunem că am măsurat valorile funcției corespun-

* Cifrele din paranteze nu reprezintă exponenți, ci numere de ordine ale punctelor, indici; asemenea notații vor fi utilizate și ulterior.

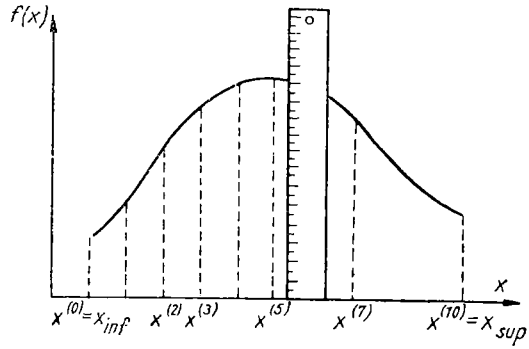


Fig. 3.2

zătoare variantelor atașate nodurilor rețelei. Să unim prin segmente rectilinii punctele de pe figură care corespund diferitelor valori ale funcției. O asemenea construcție se numește de obicei *interpolare liniară*. Rezultatul va fi o funcție nouă, f_l, al cărei grafic va fi o linie frîntă. În punctele rețelei, valorile $f_l(x)$ ale acestei funcții coincid cu valorile funcției $f(x)$ al cărei maxim îl căutăm, dar în punctele intermediare aceste două funcții probabil că se vor deosebi. Spunem „probabil", întrucît de fapt nu știm care este comportarea funcției $f(x)$, adică cunoaștem valorile $f(x)$ în punctele de diviziune, dar nu cunoaștem toate valorile $f(x)$.

Dar acum este cazul să ne amintim că funcția $f(x)$ este lină; or, dacă este lină, valorile ei în punctele intermediare nu pot să se schimbe brusc. Prin urmare, este probabil ca diferențele față de variația cea mai lină — variația rectilinie — să fie mici și într-o anumită măsură neglijabile, deci căutarea maximului funcției $f(x)$ poate fi înlocuită prin căutarea maximului funcției $f_l(x)$, ceea ce este cu mult mai simplu. Se vede de îndată că funcția $f_l(x)$ are valoarea cea mai mare în punctul de pe segment pentru care este maximă valoarea pe rețea. Astfel, am redus problema căutării pentru un număr infinit de variante la problema căutării (riguros vorbind: numai a soluției

aproximative) pe un număr finit de variante, și anume numărul punctelor de pe rețea. Alegînd din aceste variante pe cea mai bună, adică pe cea care corespunde celei mai mari dintre valorile măsurate, vom obține răspunsul aproximativ la întrebarea inițială: care din mulțimea infinită a variantelor inițiale este cea mai bună. În cazul exemplului ilustrat de figura 3.2, aceasta va fi varianta corespunzătoare numărului $x^{(5)}$.

Să observăm acum că, formal, pentru o atare simplificare nu exista necesitatea de a construi funcția $f_l(x)$, aproximativ asemănătoare cu funcția nu deplin cunoscută $f(x)$. Ea ne-a fost necesară numai pentru a înțelege fondul chestiunii. Această lămurire se putea face și cu ajutorul altor funcții apropiate de $f(x)$ și egale cu ea în punctele rețelei. Exemplul cel mai simplu îl ilustrează figura 3.3, unde funcția inițială este înlocuită cu funcția constantă pe porțiuni $f_c(x)$. Valorile noii funcții interpolatoare rămîn neschimbate în cazul unei deplasări de cel mult jumătate pas la stînga sau la dreapta nodului.

Bazîndu-ne pe ipoteza că funcția este lină, putem afirma că și maximul funcției $f_c(x)$ nu diferă decît puțin de cel al funcției inițiale; or, maximul funcției interpolatoare este evident cea mai mare dintre valorile de pe rețea, $f(x^{(5)})$. Rezultatul este același ca în cazul interpolării liniare, dar cu o mică deosebire, care este esențială. Maximul $f_c(x)$ nu este atins numai într-un singur punct. În intervalul de la $x^{(5)} - \Delta/2$ pînă la $x^{(5)} + \Delta/2$ valoarea $f_c(x)$

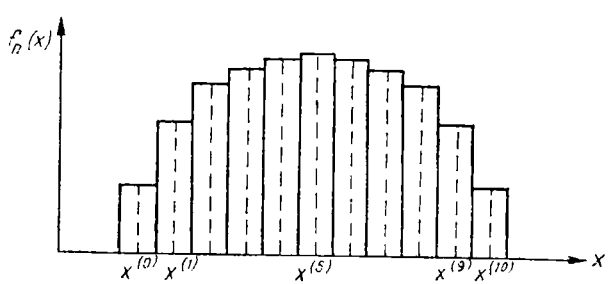


Fig. 3.3

nu se schimbă, prin urmare orice valori ale lui x, orice variante care se află între acestea sînt echivalente și fiecare dintre ele se poate lua cu aproximație drept optimă. Se pare că o asemenea concluzie ar trebui considerată mai onestă decît concluzia cu privire la caracterul optim al unei singure variante, întrucît, în mod evident, ideea de „lin" de care am făcut uz nu exclude nici una dintre posibilitățile de comportare între nodurile rețelei care sînt ilustrate în figura 3.4.

Să tragem concluziile examinării cazului unidimensional. Ne-am lămurit că **dacă ținem seama de ipoteza că funcția este lină, putem să înlocuim funcția respectivă cu una apropiată, interpolînd un număr finit de valori în nodurile rețelei;** *pentru funcția interpolatoare găsirea variantei optime, sau cel puțin a unui mic interval în interiorul căruia se află această variantă, devine o treabă relativ ușoară: este necesar doar să comparăm valorile din noduri și să alegem pe cea mai mare dintre ele.*

Dar la fel putem proceda și în probleme mult mai dificile, multidimensionale, în probleme de căutare a extremumului la funcțiile de mai multe variabile.

După cum am mai spus, în cazul bidimensional, mulțimii variantelor îi corespunde mulțimea punctelor din planul axelor de coordonate, unde pe axa absciselor marcăm valorile uneia dintre variabile, iar pe axa ordonatelor valorile celeilalte. Să notăm aceste variabile cu x_1 și x_2. În problema regimului tehnologic optim de care am vorbit

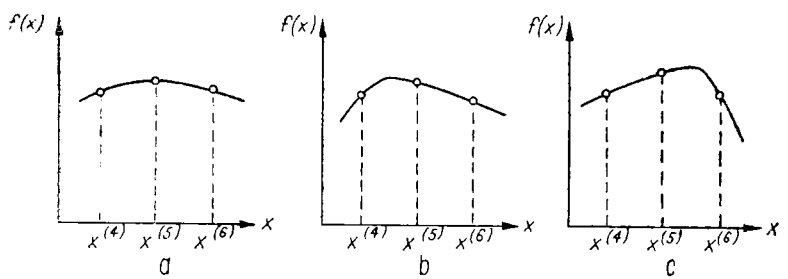


Fig. 3.4

mai sus, rolul lui x_1 îl avea viteza de amestecare v, iar rolul lui x_2 temperatura t. În această problemă, variabilele t și v puteau să varieze numai în anumite limite, în domeniul haşurat de pe figura 2.4. O asemenea situație este uzuală; vom considera că în orice problemă de căutare bidimensională variabilele x_1 și x_2 pot varia doar în limitele unui anumit domeniu mărginit, de exemplu ca cel arătat în figura 3.5. Aici variabila x_1 variază de la x_{1inf} pînă la x_{1sup}, iar x_2 de la x_{2inf} pînă la x_{2sup}. Să dividem segmentul de la x_{1inf} la x_{1sup} în m_1 părți egale, iar segmentul de la x_{2inf} la x_{2sup} în m_2 părți egale (în fig. 3.5 $m_1 = 11$, $m_2 = 9$). Prin punctele de diviziune ale primului segment, situat pe axa absciselor, să ducem drepte paralele cu axa ordonatelor, iar prin punctele de diviziune ale segmentului de pe axa ordonatelor să ducem drepte paralele cu axa absciselor. Atunci cele două familii de drepte, intersectîndu-se, vor forma o rețea; astfel, în cazul bidimensional, denumirea de „rețea", ca și denumirea de „nod al rețelei" atribuită punctelor de intersecție, apar cît se poate de firești.

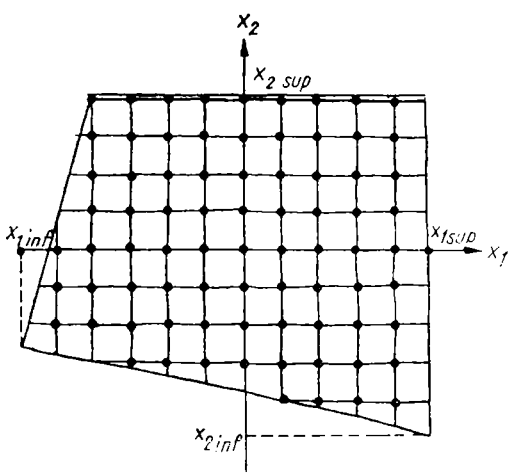


Fig. 3.5

Rețeaua considerată are pasul

$$\Delta_1 = \frac{x_{1sup} - x_{1inf}}{m_1}$$

relativ la coordonata x_1 și pasul

$$\Delta_2 = \frac{x_{2sup} - x_{2inf}}{m_2}$$

relativ la coordonata x_2; numărul total al nodurilor de pe rețea (în limitele domeniului variației posibile din fig. 3.5) este întrucîtva mai mare decît produsul $m_1 m_2$; în cazul unui domeniu dreptunghiular, acest număr este $(m_1 + 1)(m_2 + 1)$.

Pînă acum, toate operațiile noastre de construire a rețelei au fost formale, deci se puteau face pe hîrtie. Să trecem acum la acțiune; să măsurăm valorile funcției pentru valorile variabilelor egale cu valorile coordonatelor în punctele rețelei. Să construim mintal funcția de interpolare f_c, ale cărei valori în nodurile rețelei coincid cu valorile măsurate $f(x_1, x_2)$ și rămîn egale cu acestea în cazul unei deplasări de cel mult un semipas, într-o direcție sau alta pornind de la noi. Întrucît comportarea funcției f, deși necunoscută, este în concordanță cu bunele maniere, funcția de interpolare trebuie să fie apropiată de ea în toate punctele domeniului variațiilor posibile și prin urmare maximul $f(x_1, x_2)$ va fi apropiat de maximul $f_c(x_1, x_2)$. Or, este evident că acesta din urmă se poate determina prin examinarea directă a tuturor valorilor măsurate și selectarea celei mai mari.

După cum vedeți, în cazul bidimensional, raționamentul nu se deosebește prin nimic de cel din cazul unidimensional. La fel putem raționa și în cazul tridimensional și în orice alt caz multidimensional. *Deși este imposibil să ne închipuim că am putea privi o rețea în cazul cvadridimensional, modul de construire a coordonatelor nodurilor este la fel de clar ca și mai înainte; intervalul variațiilor posibile ale lui* x_1 *îl dividem în* m_1 *părți, pe cel al variabilei* x_2 *în* m_2 *părți și așa mai departe, iar apoi, măsurînd valorile*

funcției în toate combinațiile posibile ale valorilor fiecărei variabile corespunzătoare punctelor de diviziune, alegem pe cea mai mare.

Pentru amatorii de formule toate acestea se pot scrie în modul următor: determinăm cea mai mare dintre valorile funcției f de n argumente

$$f(x_1, x_2, ..., x_n)$$

pentru toate combinațiile următoarelor valori ale coordonatelor:

$$x_i = x_{iinf} + k_i\Delta_i$$

unde

$$\Delta_i = \frac{1}{m_i}(x_{isup} - x_{iinf}), \text{ cu } i = 1, ..., n;\ k_i = 0,1, ..., m_i.$$

Toate acestea sînt foarte simple, dar... destul de obositoare. Într-adevăr, numărul măsurătorilor și comparațiilor în cazul cvadridimensional este aproximativ egal cu produsul $m_1m_2m_3m_4$. Dacă pentru toate variabilele, segmentele variației posibile sînt divizate în același număr de m părți, atunci în cazul n-dimensional numărul total al nodurilor va fi m^n. Or, numărul n al variabilelor (dimensiunea problemei) poate fi mult mai mare decît o sută (în cele ce urmează cititorul se va convinge că asemenea probleme se întîlnesc destul de des, cel puțin mai des decît ar fi de dorit). Iar dacă $n = 100$ și pentru fiecare variabilă intervalul de variație este divizat numai în două porțiuni, atunci numărul nodurilor este 3^{100}, adică aproximativ 10^{48}, ceea ce înseamnă 1 urmat de 48 zerouri, un număr despre care cu greu s-ar putea spune că inspiră stimă; mai curînd, panică.

Să ne închipuim acum că valorile funcției nu se măsoară, ci doar se calculează. Să presupunem că pentru acest calcul sînt necesare o sută de operații elementare. Numărul total al operațiilor elementare pe care le pot efectua toate calculatoarele în timp de o secundă nu depășește 10^{9}*. Prin urmare, mașinile pot calcula într-o secundă nu mai mult decît 10^7 valori ale funcției, iar într-un an aproxi-

* Autorul se referă la situația din anul 1970 (*N.T.*).

mativ $3 \cdot 10^{14}$. Așadar, pentru 10^{48} valori va fi nevoie de 10^{33} ani. Este evident că astrul solar care dă viață întregii omeniri va reuși să se stingă înainte ca noi să obținem rezolvarea problemei care ne interesează! Dar poate că acestea sînt „dificultăți temporare" datorite stadiului de dezvoltare încă imperfectă a tehnicii de calcul de care dispunem astăzi? Este ușor să ne convingem de contrariul. Realizarea oricărei operații pretinde obligatoriu realizarea unui fenomen fizic. Cele mai scurte ca durată dintre fenomenele fizice cunoscute, reacțiile nucleare care au loc în cazul „interacțiunilor tari", durează aproximativ 10^{-23} secunde, iar unitatea de lungime tipică pentru mecanica cuantică (care este mecanica microcosmosului) este așa-numita lungime *compton*, egală cu $3{,}9 \cdot 10^{-11}$ cm, în timp ce unitatea de timp este 10^{-21} secunde, adică intervalul de timp necesar unei raze de lumină pentru a parcurge o lungime de un compton. Să presupunem că fenomenele de dimensiuni cuantice ar fi puse în serviciul tehnicii de calcul. Atunci viteza mașinii ar putea atinge 10^{21} pînă la 10^{23} operații pe secundă, adică 10^{28} pînă la 10^{30} operații într-un an. Dar... realizarea a 10^{50} operații va rămîne și în acest caz o cauză pierdută anticipat.

„Ei bine, va spune în concluzie cititorul chinuit de zecile de zerouri, dacă ați pornit de la probleme disperat de complicate, este firesc să ajungeți la rezultate care interzic orice speranță".

Află însă, dragă cititorule, că problema căutării maximului unei funcții depinzînd de o sută sau chiar de o mie de variabile nu este nicidecum disperată. În multe cazuri ea poate fi rezolvată și se rezolvă cu succes în practică chiar cu ajutorul mașinilor atît de lente de care dispunem astăzi.

Dar pentru ca aceasta să devină posibil, funcția trebuie să aibă o purtare și mai cuviincioasă, și mai manierată decît am presupus înainte, sau, mai precis, decît funcția la care am făcut apel înainte.

Studiul unor atare funcții și utilizarea proprietăților lor vor fi obiectul capitolului următor; deocamdată să mai reflectăm la cele aflate.

Formal, schema căutării se reducea la următoarele: la început se planifica colecția variantelor posibile (nodurile rețelei), apoi se măsurau rezultatele realizării lor (valorile funcției în noduri); din aceste valori se alegea cea mai mare (dacă se căuta maximul) sau cea mai mică (dacă se căuta minimul) și tocmai această valoare era acceptată drept maximul, respectiv minimul căutat. Drept aproximativ optime erau considerate acele variante ale căror coordonate diferau cu cel mult un semipas al rețelei de coordonatele nodurilor unde se constatase valoarea optimă.

La baza unei asemenea scheme se afla în principiu **ipoteza potrivit căreia în cazul unui pas suficient de mic o funcție lină nu se poate abate mult de funcția care o interpolează și cu care coincide în noduri.**

Dar ce anume pas poate fi considerat „mic" — aceasta depinde de gradul de regularitate, mai precis de mărimea celei mai mari variații posibile a funcției între noduri. Dacă această variație maximă nu este cunoscută dinainte, ea poate fi practic estimată după variația valorilor pe rețea. Întrucît coordonatele nodurilor au fost alese independent de orice proprietăți ale funcției, se poate presupune că diferența dintre valorile în noduri și valorile intermediare nu depășește ca mărime cea mai mare dintre diferențele între două noduri vecine ale rețelei, ceea ce ne va oferi valoarea erorii în determinarea extremumului. Desigur, această eroare se poate determina numai după ce executăm măsurătorile pe rețeaua dată.

Dar ce ne facem dacă eroarea va fi prea mare? N-avem decît, firește, să facem rețeaua mai densă, recurgînd la un număr mai mare de măsurători*.

Așadar, putem estima eficiența alegerii rețelei, gradul „micimii" pasului numai după executarea măsurătorilor

---

* Evident, practica aceasta este indicată numai pentru un domeniu de variație în care parametrii se schimbă mult.

pe rețeaua respectivă și după prelucrarea rezultatelor, adică după studierea funcției înseși.

Ideea de funcție continuă, lină, regulată ne-a sugerat posibilitatea înlocuirii unui număr infinit de variante cu un număr finit, dar această idee era lipsită de un caracter suficient de determinat pentru ca noi să ne putem da seama care anume număr de variante este suficient pentru estimarea optimului cu aproximația dorită.

O clasă suficient de întinsă de funcții satisface condiția cunoscută în matematică sub numele de *condiția lui Lipschitz:* pentru orice puncte $x^{(1)}$ și $x^{(2)}$ există o constantă L, astfel încît

$$|f(x^{(1)}) - f(x^{(2)})| \leqslant L\,|x^{(1)} - x^{(2)}|.$$

Atunci diferența dintre valorile funcției în orice punct intermediar și valorile din noduri nu depășește $L\Delta$, adică constanta lui Lipschitz este tocmai acea măsură numerică a regularității a cărei cunoaștere este suficientă pentru a alege mărimea pasului în cazul unui grad de precizie dat. De fapt, procedura practică de estimare a erorii despre care am vorbit mai sus se reduce la estimarea constantei lui Lipschitz pe baza măsurătorilor în punctele rețelei realizate.

4
Lupta corp la corp

(Căutarea locală pe rețea)

Pitică iarbă, pașii ți-s mărunți,
dar ții sub talpă-ntreg pămîntul!
Rabindranath Tagore, *Păsări răzlețe* *

Înspăimîntătoarele numere la care am ajuns planificînd căutarea pe o rețea dată în prealabil ne determină să căutăm alte idei, ne constrîng să recurgem la altă procedură de planificare a experimentelor pentru verificarea eficienței variantelor. Este însă foarte greu să născocim ceva nou, căci nu știm dinainte nimic despre comportarea funcției, în afară de faptul că ea este relativ regulată. Orice variantă, orice punct din domeniul admis poate fi bănuit în aceeași măsură că ar fi punctul optim. În acest caz, ce plan ar putea să fie mai rezonabil decît așezarea echidistantă a punctelor de cercetare — nodurile rețelei? Într-adevăr, un plan mai bun, construit dinainte, e cu neputință de imaginat. Înseamnă, așadar, că nu există ieșire? Înseamnă oare că va trebui să ne împăcăm cu ideea că omenirea nu va reuși să rezolve probleme atît de interesante cum sînt cele relative la optimizarea unei funcții care nu are, la urma urmelor, decît o sută de variabile?

---

* **R. Tagore,** *Poeme*, trad. de George Dan, Editura pentru literatura universală, București, 1967 (*N.T.*).

Din fericire, o soluție există, iar răspunsul este tot atît de paradoxal pe cît de instructiv: trebuie să renunțăm la trasarea în prealabil a planului de acțiune. *Planul trasat dinainte, programul rigid de așezare a punctelor de cercetare trebuie să fie înlocuit cu un plan suplu, în cadrul căruia alegerea fiecărei variante supuse verificării depinde de rezultatele obținute la testarea variantelor anterioare.* Să trecem la înfăptuirea acestei noi idei, dar cu prudență, menținînd la început unele trăsături ale principiului vechi: trierea valorilor pe rețeaua dată în prealabil.

Să considerăm o problemă bidimensională în care mulțimea variantelor este reprezentată de mulțimea punctelor din dreptunghiul arătat în figura 4.1. Așezăm pe dreptunghi o rețea cu 35 de noduri. La început facem uz de metoda trierii totale. Pentru aceasta va fi nevoie să facem măsurători în toate punctele rețelei, comparînd consecutiv valorile măsurate și alegînd valoarea minimă (pentru variație, presupunem că varianta optimă este aici acea variantă care dă valoarea minimă). Ordinea, adică succesiunea variantelor alese pentru măsurare, este indiferentă, dar o anumită ordine trebuie adoptată,

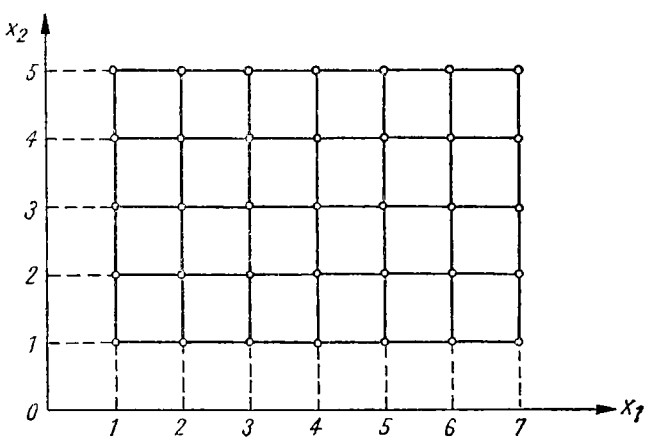


Fig. 4.1

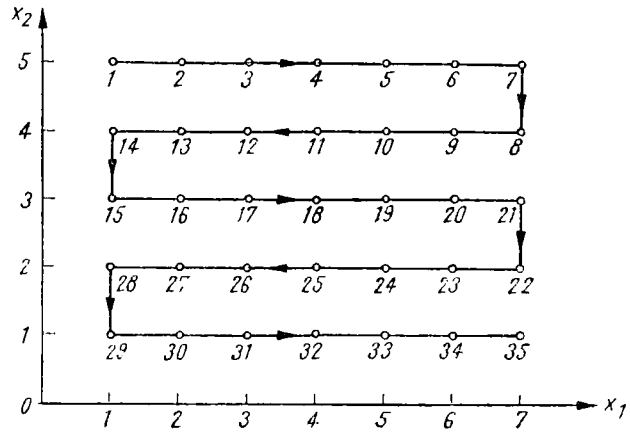


Fig. 4.2

pentru a nu scăpa din vedere vreun punct. Cel mai frecvent se face uz de două metode: măturarea rînd cu rînd (fig. 4.2) și desfășurarea în spirală (fig. 4.3). Cuvîntul „desfășurare" se folosește pentru că succesiunea variantelor examinate pare a se desfășura în lanț unidimensional (în timp), iar cuvîntul „spirală" corespunde aspectului drumului pe care îl străbatem în plan trecînd de la o variantă la alta (în figuri, variantele sînt numerotate în ordinea măsurătorilor executate).

Așadar, acționînd într-o ordine dinainte stabilită, am parcurs toate punctele rețelei, am executat măsurătorile (cercurile din figura 4.4 indică rezultatele corespunzătoare) și am stabilit că minimul se află în punctul cu coordonatele $x_1 = 5$, $x_2 = 4$ (sau mai precis în vecinătatea sa). Pentru aceasta, repet, au fost necesare 35 de măsurători. Acum să uităm această cale greoaie și să încercăm să realizăm căutarea într-un mod nou, ținînd seama de ideea nouă conform căreia „alegerea unei variante date trebuie să depindă de rezultatele testării variantelor precedente".

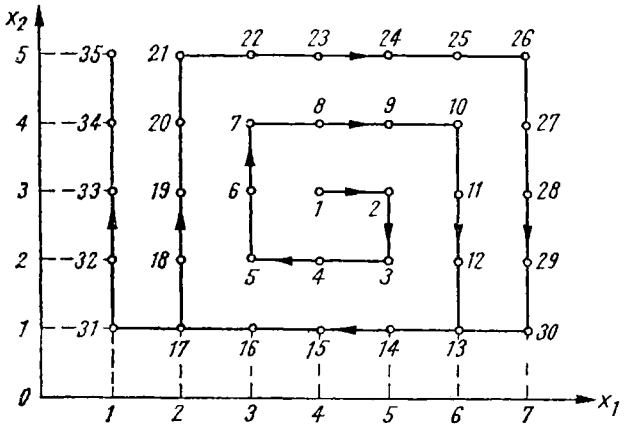

x_2
x_1
0
1
2
3
4
5
6
7
−35
−34
−33
−32
−31
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

Fig. 4.3

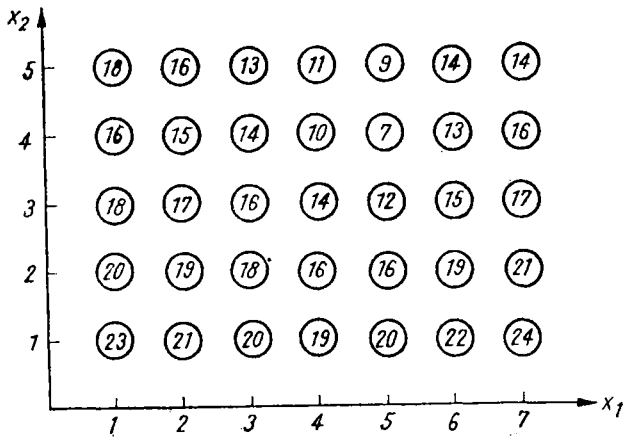

x_2
x_1
1
2
3
4
5
6
7
18 16 13 11 9 14 14
16 15 14 10 7 13 16
18 17 16 14 12 15 17
20 19 18 16 16 19 21
23 21 20 19 20 22 24

Fig. 4.4

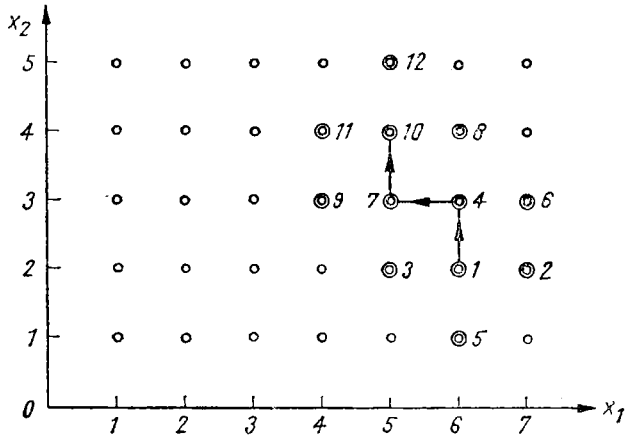


Fig. 4.5

Varianta inițială nu are o precedentă, deci putem să începem cu oricare variantă, pe care o alegem în mod arbitrar dintre cele posibile, considerînd drept posibile, deocamdată, numai acele variante care corespund nodurilor așezate anterior pe aceeași rețea. Fie acest prim nod punctul de coordonate $x_1 = 6$, $x_2 = 2$. Vom atribui acestui punct numărul 1 (fig. 4.5). Prima măsurătoare dă valoarea funcției $f^{(1)} = 19$, rezultat care, firește, este același ca și în vechea schemă de măsurători.

Să fi nimerit oare, întîmplător, cea mai bună variantă, care ne dă valoarea minimă a funcției? Dacă este așa, atunci ea, fiind mai bună dintre toate, este mai bună decît vecinele ei de pe rețea. Să verificăm acest lucru: executînd cîte un pas la dreapta și la stînga, în sus și în jos, facem alte patru măsurători și comparăm rezultatele*. Constatăm că variantele 3 și 4 sînt mai bune decît cea inițială și că degeaba am bănuit că varianta 1 ar fi optimă. Acum cea mai bună dintre variantele testate, varianta 4, dă

* Dacă nodul bănuit se află pe frontieră, atunci se cercetează numai „vecinii" din interiorul domeniului admis.

valoarea $f^{(4)} = 15$. Să fie aceasta cea mai bună? Iarăși revenim la vecini: deplasarea la dreapta și la stînga, în sus și în jos va cere trei noi măsurători. Și iarăși rezultatele arată: varianta 4 nu este cea mai bună deoarece în punctul 7 avem $f^{(7)} = 12$. Trecem acum la o nouă bănuială, de astă dată în ce privește varianta 7; iarăși încercarea vecinilor, dintre care doi sînt noi, netestați. Acum bănuiala se va referi la unul din ei, căruia i s-a atribuit numărul 10: $f^{(10)} = 7$. Încercăm vecinii și noile puncte 11 și 12 declară într-un glas: „Nu, noi nu sîntem mai bune decît punctul 10!" Așadar, varianta 10 cu valoarea sa $f^{(10)} = 7$ este recunoscută drept cea mai bună de către toate vecinele ei. (Dacă ancheta s-ar referi la vecini de locuință, atunci probabil că niciodată n-am obține o asemenea unanimitate, dar matematica este tare — sau poate slabă? — tocmai prin aceea că înlocuiește toate emoțiile în aprecierea a ceea ce este „mai bun" prin limbajul simplu al cifrelor.) Noi, care anterior am realizat munca grea de testare a tuturor celor 35 de variante, știm că în cazul dat „vocea vecinilor este vocea Domnului": varianta 10 avînd coordonatele (5, 4) este realmente mai bună decît toate. Căutarea sa a pretins numai 12 măsurători, dintre care „bănuite de a fi optime" au fost, pe traseu (traseul este indicat în figura 4.5 prin săgeți) numai patru. Dar poate că aceasta este doar o întîmplare, un rezultat al alegerii convenabile a punctului inițial? Nu, lucrurile nu stau așa. Cititorul poate verifica că chiar dacă varianta inițială este aleasă în modul cel mai nefericit — punctul avînd coordonatele (2, 1) — numărul total al măsurătorilor este 21, adică mult mai mic decît numărul total al nodurilor rețelei.

Astfel, am găsit un mijloc de a înlătura necesitatea trierii totale, iar acest mijloc este deosebit de simplu: căutarea începe cu un nod ales în mod arbitrar; se fac măsurători în nodurile apropiate; este ales dintre acestea cel care dă rezultatul cel mai bun, se estimează rezultatele în nodurile apropiate de cel ales etc. Procedurile se repetă în mod identic (ciclic) pînă în momentul cînd se va găsi o variantă mai bună decît toate care îi sînt apropiate. Această variantă va fi tocmai cea optimă. Procedurile de

asemenea gen (iar ele sînt foarte variate, după cum ne vom convinge în curînd) se numesc *proceduri iterate de căutare locală.* „Iterat" înseamnă repetat, iar cuvîntul „local" indică faptul că se compară numai vecinii.

Meritele acestei scheme se vor dovedi și mai mari dacă vom micșora pasul rețelei. Din figura 4.6 este clar că reducînd pasul rețelei de două ori (după fiecare coordonată în parte) numărul total al nodurilor crește aproape de patru ori: $13 \times 9 = 117$; numărul măsurătorilor necesare se mărește mai puțin decît de două ori. (În figura 4.6 se arată rezultatele măsurătorilor exact în punctele traseului către optim; numărul total al măsurătorilor este douăzeci.) Avantajul trecerii la schema de căutare iterată devine și mai mare în problemele multidimensionale. Păcat că nu vom putea reprezenta aceasta pe o imagine intuitivă, dar cauza succesului trebuie să fie clară: realizînd căutarea iterată ne mișcăm de-a lungul unor curbe, denumite *traiectoriile căutării* (tocmai o asemenea traiectorie o arată săgețile din fig. 4.5).

Dacă aceste traiectorii nu sînt prea complicate, numărul punctelor de pe ele nu depășește prea mult numărul punc-

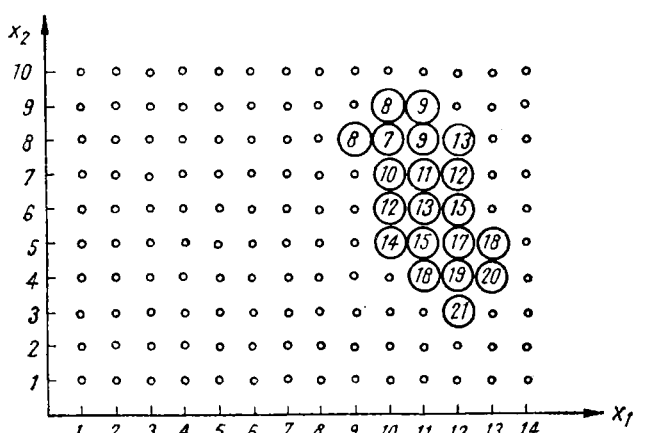


Fig. 4.6

telor din rețelele unidimensionale în care sînt divizate domeniile de variație ale fiecărei coordonate în parte. În cap. 3 am notat aceste numere cu $m_1 + 1$, $m_2 + 1$, ..., astfel încît lungimea traiectoriei căutării poate fi estimată prin numărul de pași.

$$m_1 + 1 \text{ în cazul unidimensional}$$

$$m_1 + m_2 + 2 \text{ în cazul bidimensional}$$

$$m_1 + m_2 + m_3 + 3 \text{ în cazul tridimensional etc.}$$

Dacă toate acestea sînt egale, atunci numărul punctelor traiectoriei în cazul n-dimensional poate fi apreciat prin numărul nm (în toate exemplele considerate de noi traiectoriile au fost mult mai scurte). Să observăm că în mișcarea noastră trebuie să-i cercetăm și pe vecini, nodurile apropiate de nodurile aflate pe traiectorie. Asemenea vecini în cazul unidimensional sînt doi pentru fiecare punct; în cazul bidimensional ei sînt $4 = 2 \times 2$, în cazul tridimensional, $6 = 2 \times 3$, iar în cazul n-dimensional numărul vecinilor fiecărui punct este $2n$. (De aici ar trebui scăzuți cei doi vecini care se află pe traiectorie, dar pentru un n mare numerele $2n$ și $2n - 2$ diferă prea puțin unul de altul.) Așadar, numărul total al măsurătorilor necesare pentru realizarea căutării iterate poate fi apreciat ca fiind numărul $2n^2m$. Să vedem ce se întîmplă în cazul înspăimîntător a o sută de variabile ($n = 100$) cu rețeaua $m = 10$ Estimația noastră (foarte grosolană!) dă doar $2 \cdot 10^5$ variante, ceea ce nu se poate compara cu numărul gigantic, avînd o sută de zerouri, care determină numărul total al variantelor ce ar fi trebuit cercetate în cazul trierii complete!

Așadar, ne putem mîndri: am obținut un succes decisiv, vom putea să facem față problemelor celor mai înfricoșătoare, aducînd truda noastră în dar omenirii, ba chiar, dîndu-ne modest la o parte, putem să așteptăm ca omenirea emoționată și recunoscătoare să ne acorde laurii

învingătorilor. Și totuși... se cuvine să ne ferim de înfumurare și să reflectăm cu ce preț s-a obținut succesul și care este valoarea reală a acestui succes.

În realitate, acest succes are la bază faptul că noi am trădat simpla și nobila idee a continuității funcției, pe care se întemeiază trierea totală pe rețea și am înlocuit-o cu presupunerea înspăimîntător de îngustă ca orizont că orice variantă mai bună decît cele vecine este mai bună decît toate celelalte. Cu alte cuvinte, **am înlocuit problema căutării extremumului global prin problema cu mult mai simplă a căutării extremumului local.** *Or, soluțiile acestor probleme coincid numai în două cazuri: dacă funcția este astfel încît în întregul domeniu de variație a variabilelor ea are un singur extremum local, sau dacă întîmplător extremumul local de care am dat este tocmai cel global.*

Acestea sînt sacrificiile la care am consimțit: am pîngărit în mod grosolan neprihănitul altar al generalității. Dar... așa cum nu se cuvenea să ne facem o părere prea măgulitoare despre noi înșine, tot astfel nu trebuie să ne vedem nici în culori prea negre: dacă am fi rămas credincioși ideii pure de continuitate, am fi fost pur și simplu incapabili să abordăm problemele multidimensionale; sacrificînd însă această idee, am obținut o șansă dacă nu pentru extremum extremorum, cel puțin pentru un anumit extremum. În afară de aceasta, există o mulțime de funcții despre care se știe bine că nu posedă decît un extremum local.

Să observăm, în primul rînd, că, dacă o funcție este cu adevărat regulată, adică variază nu numai puțin, dar și treptat, lin, atunci, descrescînd în apropierea unui maxim local, ea nu-și va modifica comportarea nici în cazul unei oarecare îndepărtări de acest maxim, adică: dacă domeniul căutării conține în interiorul său un extremum local și dacă întinderea acestui domeniu nu este prea mare (noțiune care nu este destul de bine definită, observație la fel de adevărată pentru noțiunile de „regularitate", „bună creștere" și altele asemănătoare), atunci acest extremum va fi aproape sigur unicul.

Astfel, trecînd de la trierea completă pe rețea la procedurile de căutare locală iterată, putem să ne justificăm conduita prin aceeași referire la ipoteza regularității funcției, ce-i drept, dînd de astă dată acestei noțiuni un sens mult mai restrîns. Pentru a ne convinge definitiv de caracterul restrictiv al noii ipoteze, să privim figura 4.7, care reprezintă o funcție întru totul regulată, de un singur argument. Îi vom căuta minimul în domeniul admis, mărginit de numerele 2 și 10. Să construim rețeaua cu pasul $\Delta = 1$ și să trecem la realizarea procedurii de căutare locală. Nu e greu de văzut că, indiferent de nodul de la care am pornit căutarea, aceasta se va încheia în nodul $x = 5$, $f = 2{,}1$. Evident, același punct îl vom obține și prin trierea directă. Aici minimul local al funcției este și minim global. Să încercăm însă acum să căutăm o variantă corespunzătoare nu valorii celei mai mici a funcției, ci valorii celei mai mari. Să începem, de exemplu, cu punctul $x = 4$, în care avem $f = 2{,}2$. Verificăm vecinii: pentru $x = 3$ aflăm $f = 2{,}4$; pentru $x = 5$ aflăm $f = 2{,}1$. Cu aceasta, bănuiala se transferă asupra punctului $x = 3$. Repetînd procedura, ajungem în punctul $x = 2$, unde $f = 3$. Mai departe n-avem încotro merge, deoarece în domeniul admis acest punct are un singur vecin, $x = 3$,

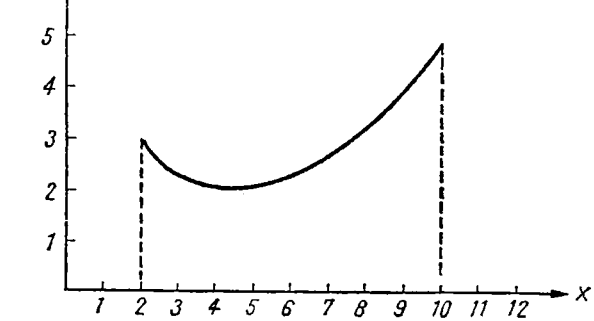


Fig. 4.7

pentru care valoarea funcției este mai mică. Astfel, pentru $x = 2$ am găsit maximul. Dar noi, care ne-am așezat „mai sus" și privim de acolo întreaga variație a valorilor funcției, înțelegem că maximul găsit nu este global, întrucît maximul global este egal cu 5 și corespunde variantei $x = 10$.

Așadar, căutarea locală poate să nu conducă la un succes deplin chiar și în cazul unei funcții cu o atît de aleasă bună creștere.

Ne putem consola întrucîtva cu faptul că, dacă punctul inițial ar fi fost undeva la dreapta punctului $x = 4$, am fi ajuns la o soluție corectă. În cazul bidimensional, lucrurile stau și mai prost, de exemplu în cazul funcției pentru care figura 4.4 indică valorile pe rețea, minimul local, după cum ne-am convins, este unic, dar maximele locale sînt în număr de trei; căutarea locală poate să conducă la oricare dintre ele, aceasta depinzînd de varianta pe care o luăm drept inițială.

Așadar ne-am convins definitiv că procedurile de căutare locală iterată au posibilități restrînse; continuăm să susținem totuși că ele reprezintă o mare invenție. Nimeni nu ne va putea acuza de lipsă de modestie, deoarece este binecunoscut că printre inventatorii reali ai procedurilor locale se numără asemenea mari savanți ca Newton și Gauss. Ei erau nu numai matematicieni, ci și cercetători ai naturii și înțelegeau de minune întreaga importanță practică a rezolvării problemelor de optim. E drept că noi am făcut cunoștință doar cu cea mai primitivă dintre procedurile de iterare cunoscute. Probabil că atît Newton cît și Gauss ar fi respins în mod hotărît propunerea ca schema de „verificare a vecinilor" descrisă mai sus să fie asociată cu numele lor, dar totuși și în această schemă au ieșit la iveală unele proprietăți ale procedurilor de căutare mai elegante și mai eficiente, proceduri la a căror descriere vom trece în cele ce urmează.

5
Gîndire și acțiune

(Căutarea proporțională)

După codița unei flori,
spunea marele Bacon, putem
afla direcția vîntului
Byron, *Don Juan*

Orice inovație apare în urma răzvrătirii împotriva tradițiilor statornicite, dar orice lucru nou poartă întotdeauna și semnele vechiului. Astfel, procedura de iterare pe care am descris-o amintește tot atît de puternic de proprietățile metodei trierii după nodurile rețelei, cum forma primelor automobile imită pe strămoșul lor, trăsura cu cal. Într-adevăr, dacă nu avem de gînd să triem toate variantele de pe rețea, să programăm dinainte întreaga succesiune a experimentelor și dacă ne limităm doar la căutarea unui extremum local, de ce să mai menținem rețeaua însăși? Să renunțăm deci și la ea, păstrînd din procedura de iterare pe care o cunoaștem numai ceea ce este nou: ideea de a alege varianta curentă pe baza comparării variantei anterioare cu vecinii ei; și să mai introducem un element nou, noțiunea de *previziune*, de *prognoză* a comportării funcției. Să vorbim mai amănunțit despre aceasta, mărginindu-ne deocamdată numai la exemplele cele mai simple, problemele unidimensionale.

Să presupunem că este necesar să căutăm maximul funcției f de un argument, fără a cunoaște alura acestei funcții și dispunînd doar de posibilitatea de a măsura sau calcula valorile funcției în diferite puncte. Așa stînd

lucrurile, alegem un punct arbitrar $x^{(0)}$ și aflăm $f(x^{(0)})$. Pe baza unei singure măsurători nu putem judeca nimic, nu putem spune dacă alegerea inițială e bună sau rea. Pentru a putea judeca trebuie să facem cel puțin încă o măsurătoare și să comparăm apoi rezultatele. Pentru aceasta vom alege un punct apropiat avînd coordonata $x^{(0)} + \Delta$, unde Δ este o mărime pozitivă mică.

Avem trei posibilități:

a) $f(x^{(0)} + \Delta) > f(x^{(0)})$, creșterea valorii funcției odată cu creșterea argumentului;

b) $f(x^{(0)} + \Delta) < f(x^{(0)})$, descreșterea valorii funcției odată cu creșterea argumentului;

c) $f(x^{(0)} + \Delta) = f(x^{(0)})$, valoarea funcției rămîne constantă.

Deocamdată să facem abstracție de ultima posibilitate, deoarece este clar că în cazul unei alegeri arbitrare a punctului inițial egalitatea este un caz de excepție, practic improbabil.

În ce privește rezultatele de tipul *a* sau *b* se pot trage unele concluzii importante. Deoarece funcția este regulată și valorile ei variază lin, tendințele de creștere sau descreștere remarcate în cazul micii deplasări Δ probabil că se vor menține și în cazul unei îndepărtări mai mari de punctul inițial.

Pentru simplitate, să presupunem că modul de comportare a funcției poate fi bine aproximat de dreapta trasată prin cele două puncte măsurate (fig. 5.1*a*, *b*). Odată adoptată această ipoteză, planul acțiunilor ulterioare este evident. În cazul *a* trebuie să facem un pas la dreapta, în direcția creșterii argumentului, iar în cazul *b* un pas la stînga, în direcția descreșterii sale, după care, conform cu principiul de bază al proceselor de iterare, să repetăm aceeași procedură în punctul nou. Singurul lucru important este să ne dăm seama de mărimea pasului ce trebuie făcut. Din capul locului este clar că dacă l-am lua egal cu Δ, atunci am ajunge iarăși la procedura simplă pe care am studiat-o: deși rețeaua nu este construită în prealabil,

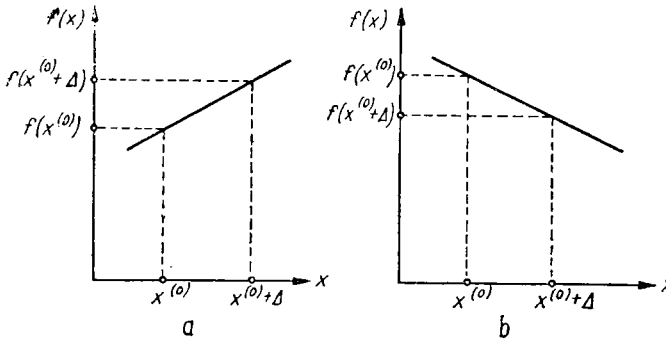


Fig. 5.1

totuși punctele de măsurat se vor situa în mod automat la distanțe egale cu Δ, astfel încît construcția se va face de la sine.

De aceea, propunem încă o ipoteză rezonabilă. Felul în care este așezată dreapta care trece prin două puncte vecine de pe graficul funcției determină înclinarea pantei în apropierea acestor puncte (fig. 5.2). Cu cît această pantă este mai înclinată, cu atît mai rapid cresc (sau descresc) valorile, așadar, cu atît mai mari șanse există pentru menținerea în continuare, eventual cu altă viteză, a creșterii, respectiv a descreșterii. Prin urmare, este rezonabil să

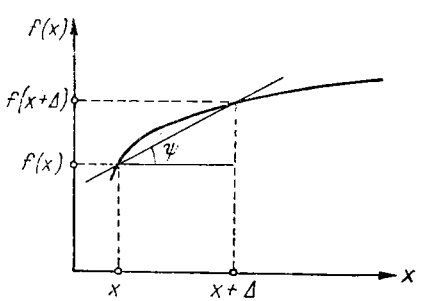


Fig. 5.2

sporim mărimea pasului odată cu creșterea înclinării pantei.

Vom estima această înclinare în apropierea punctelor x și $x + \Delta$ prin tangenta unghiului ψ pe care îl face dreapta de aproximare cu axa absciselor, adică prin raportul

$$\frac{f(x + \Delta) - f(x)}{\Delta}$$

dintre variația valorilor funcției și variația corespunzătoare a argumentului. Deoarece această mărime este diferită pentru diferiți x și Δ, o vom nota cu S (x, Δ). După cum au apărut una sau alta din cele trei situații enumerate, în punctele concrete x și $x + \Delta$, S va avea, firește, o valoare pozitivă, negativă sau nulă.

Așadar, ne-am înțeles ca mărimea pasului să depindă de mărimea înclinării calculate S; mai mult, să admitem că mărimea pasului este proporțională cu înclinarea. Atunci schema căutării iterate a maximului va fi pe deplin determinată de formulele

$$x^{(t+1)} = x^{(t)} + \theta S(x^{(t)}, \Delta);$$

$$S(x^{(t)}, \Delta) = \frac{1}{\Delta} [f(x^{(t)} + \Delta) - f(x^{(t)})]$$

unde t este numărul iterării ($t = 0, 1, 2, ...$), iar θ este un coeficient de proporționalitate pozitiv și constant. Într-adevăr, în virtutea acestei formule, din punctul inițial $x^{(0)}$ vom trece la punctul nou $x^{(1)}$, efectuînd pasul $x^{(1)} - x^{(0)} = \theta S$. Dacă $S(x^{(0)}, \Delta)$ este mai mare decît zero, adică $f(x^{(0)} + \Delta) > f(x^{(0)})$, atunci, după cum și trebuia, pasul va fi efectuat în direcția creșterii argumentului; în caz contrar, pasul va fi efectuat în direcție contrară; în ambele cazuri mărimea sa va fi proporțională cu mărimea înclinării. În punctul $x^{(1)}$ totul se va repeta: din nou va fi estimată înclinarea

$$S(x^{(1)}, \Delta) = \frac{1}{\Delta} [f(x^{(1)} + \Delta) - f(x^{(1)})],$$

ceea ce va permite aflarea mărimii și direcției noului pas ș.a.m.d.

Astfel, **întreaga procedură se desface în mod natural în iterări separate.** *În cursul fiecăreia dintre ele, la început se studiază comportarea funcției în două puncte învecinate, apoi se face uz de ipoteza că această comportare se va menține și dacă ne îndepărtăm de punctele studiate și, în fine, se execută operația, adică pasul de mărime optimă către un punct nou. În felul acesta,* **schema descrisă întruchipează unitatea dintre experiență, gîndire și acțiune.**

Este importantă că **în cadrul acestei scheme, spre deosebire de schema anterioară mai simplă, acțiunea nu este încătușată de cadrul restrîns al experienței nemijlocite: alegerea mărimii pasului se poate înfăptui într-o anumită măsură în mod independent, deoarece dispunem de coeficientul de proporționalitate θ, căruia putem să-i atribuim diferite valori.** *Acest coeficient pare să simbolizeze gradul de încredere față de diferitele ipoteze propuse și totodată gradul de nerăbdare a celui care caută.* Este evident că un coeficient mare corespunde unei încrederi mai mari și unei răbdări mai mici.

Dar care să fie gradul rezonabil de încredere, cum să dispunem de libertatea de alegere a pasului pe care ne-o oferă schema? Pentru a răspunde la această întrebare, este necesar să studiem minuțios noua schemă a căutării, care de obicei este denumită *schema proporționalității.*

Dar mai înainte trebuie să răspundem la o întrebare și mai simplă: deși cînd am construit schema porneam de la ipoteze rezonabile, *oare nu se poate întîmpla ca ea să ne conducă nu către maximul dorit (fie și local!), ci către un punct prin nimic remarcabil?*

Să ne adresăm din nou formulei de bază. Este clar că dacă înclinarea în apropierea punctului $x^{(l)}$ nu este nulă, atunci schema va pretinde în mod automat părăsirea acestui punct. Dar dacă pentru un anumit l se va dovedi că $S\ (x^{(l)},\ \Delta) = 0$, formula va da răspunsul:

$$x^{(l+1)} = x^{(l)};$$

cu alte cuvinte, nu este necesar să mai mergem mai departe. Deci punctul de posibilă oprire în cazul schemei considerate satisface condiția

$$S(x^{(l)}, \Delta) = 0$$

adică

$$f(x^{(l)} + \Delta) = f(x^{(l)}).$$

Astfel, procesul de căutare va stopa în mod automat îndată ce într-un anume punct valoarea funcției se va dovedi egală cu valoarea într-un punct întrucîtva deplasat, adică de îndată ce va apărea tocmai acea posibilitate *c* de care am făcut abstracție la începutul acestui paragraf. În momentul acela lucrul era rezonabil, întrucît acolo era vorba despre punctul inițial, în care o asemenea situație este practic improbabilă. Însă acum ne-am convins că schema iterării posedă proprietatea de a nu se astîmpăra pînă nu va parveni tocmai la o asemenea situație. Va trebui s-o examinăm mai amănunțit, să vedem cum se poate comporta funcția f în apropierea punctelor $x^{(l)}$ și $x^{(l)} + \Delta$.

În primul rînd, ținînd seama de regularitate, ne vom limita numai la trei cazuri posibile (fig. 5.3 *a*, *b*, *c*). În cazul *a*, în toate punctele dintre cele două considerate, $x^{(l)}$ și $x^{(l)} + \Delta$, valoarea funcției este mai mare decît în aceste puncte. Printre aceste puncte există unul cel mai bun x_{max}. Întrucît Δ este mic, și $f(x_{max})$ va diferi puțin de valoarea $f(x^{(l)})$ aflată în procesul căutării, astfel încît scopul, determinarea valorii optime locale, poate fi considerat atins îndată ce am aflat $x^{(l)}$. Dar la prima vedere ar fi posibil și cazul *b*; aceasta ar însemna o catastrofă, deoarece oprirea ne-ar aduce nu în vecinătatea maximului căutat, ci în aceea a unui minim. Din fericire, acest caz poate fi evitat cu ușurință. Trebuie doar să ne deplasăm puțin la stînga sau la dreapta punctului $x^{(l)}$ (fig. 5.4) și să repetăm măsurătorile, iar procedura de iterare ne va scoate în mod automat din locul periculos. Într-adevăr, să ne deplasăm întrucîtva la dreapta: atunci înclinarea măsurată va fi pozitivă, iar punctul nou aflat după for-

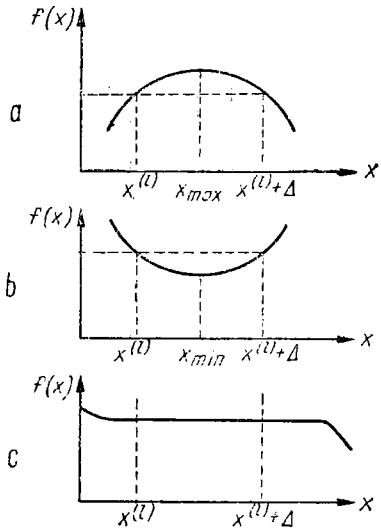


Fig. 5.3

mula căutării va fi și mai la dreapta. Dacă însă ne vom deplasa spre stînga, înclinarea va fi negativă și formula va pretinde să facem o nouă măsurătoare și mai la stînga, adică într-un punct cu o abscisă mai mică. Acum este limpede că, deși o oprire în poziția indicată de figura 5.3 *b*

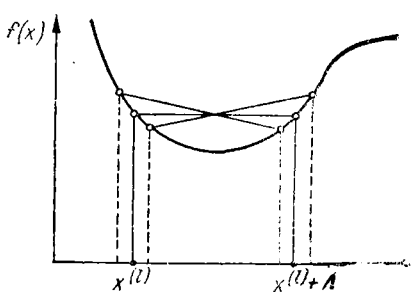


Fig. 5.4

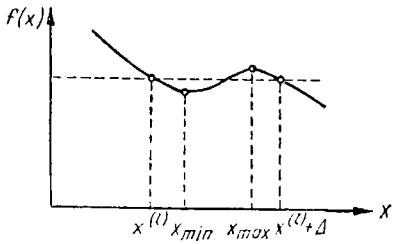


Fig. 5.5

este posibilă, totuși o asemenea situație este nestabilă față de orice schimbări mici și ca atare ea poate fi ușor evitată realizînd măsurători de control în puncte ușor deplasate față de punctele $x^{(l)}$ și $x^{(l)} + \Delta$.

În sfîrșit, este posibil și cazul reprezentat în figura 5.3 *c*, unde oprirea a avut loc nu fiindcă am fi nimerit în vecinătatea căutată a maximului și nici fiindcă ne-am rătăcit întîmplător în vecinătatea minimului. Aici punctele $x^{(l)}$ și $x^{(l)} + \Delta$ se află pur și simplu pe o dreaptă riguros orizontală.

Din punct de vedere matematic, acest caz este de-a dreptul criminal și nedorita stopare este aici inevitabilă.

Așadar, căutarea extremumului cu ajutorul metodei ce se bazează pe calculul înclinării poate să nu ducă la țel dacă pe anumite porțiuni funcția este strict constantă*.

Trebuie însă să observăm că în natură, spre deosebire de ceea ce se întîmplă în construcțiile matematice ideale, nimic nu este riguros constant.

Pentru funcția-scop, atunci cînd ea apare pe cale naturală, sînt posibile cazurile indicate în figura 5.5. Aici opririle au loc atît în vecinătatea unui maxim local cît și în vecinătatea unui minim local. Ambele sînt doar detalii mărunte pe fondul pantei fundamentale (deoarece în limitele Δ sînt imposibile variații mari), dar, *din păcate, metoda noastră este locală și ca atare nu este asigurată împo-*

* Firește că și metoda anterioară, a testării vecinilor de pe rețea, prezintă aceeași deficiență.

triva atracției exercitate de astfel de accidente ale pantei. Poate că și nereușita în cazul platformei orizontale merită să fie considerată ca un anumit caz limită al nereușitelor cu movilițe și deci considerată ca o pedeapsă inevitabilă pentru abordarea locală.

Trebuie să notăm și următoarea împrejurare, de astă dată folositoare: deși metoda este locală, totuși în cazul unor movilițe prea mici, adică al unor extreme locale neînsemnate, oprirea nu va avea loc. Ne convingem lesne de aceasta examinînd figura 5.6. Panta este plină de numeroase movilițe și gropi mici, dar ele sînt atît de neînsemnate încît funcția nu diferă de dreapta cu înclinare S cu mai mult decît mărimea β, unde $\beta < S\Delta/2$. Este clar că înclinarea calculată după valorile funcției inseși în orice puncte x și $x + \Delta$ va avea același semn ca și înclinarea S pe panta „neaccidentată". Făcînd încercări numai la distanța Δ, noi nu observăm micile movilițe de pe versant și pășim curajos în direcția corectă, deși ritmul înaintării poate fi mai mic sau mai mare decît ritmul dictat de înclinarea medie S. Dacă luăm un Δ mai mare, atunci și movilițele mai înalte vor fi omise, dar în acest caz și precizia determinării extremumului va fi mai coborîtă. Astfel încît, îngrijindu-ne de depășirea micilor movilițe, nu trebuie să uităm nici faptul că și maximul căutat nu este decît unul din extremele locale, adică tot o movilță. Dealtfel, o discuție serioasă în ce privește problema luptei cu „meschinăria" metodelor locale merită să fie amînată pînă vom ajunge la cap. 11.

Așadar analiza tuturor cazurilor posibile de oprire, de „liniștire" a procesului căutării ne-a convins că o ase-

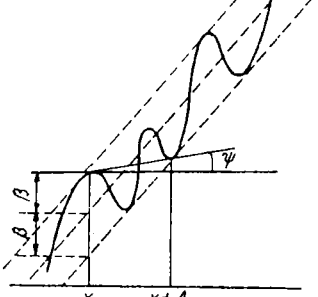


Fig. 5.6

menea oprire va avea loc numai atunci cînd undeva între punctele $x^{(l)}$ și $x^{(l)} + \Delta$ va avea loc maximul local căutat. Aceasta înseamnă că mărimea Δ, la fel ca și în schema rețelei, determină precizia definitivă a căutării. Să amintim însă că aici ritmul căutării nu depinde de Δ, ci de alegerea coeficientului θ.

Acum e și timpul să începem analiza promisă a modului în care este rezonabil să alegem acest parametru.

Să considerăm (fig. 5.7), alături de graficul funcției f, graficul funcției care se obține din aceasta printr-o deplasare a argumentului cu Δ la stînga. Este lesne să ne reprezentăm felul dependenței înclinării S de x, care diferă doar prin scară de diferența dintre cele două grafice considerate. Punctul în care se anulează $S(x, \Delta)$ este tocmai punctul $x^{(l)}$ al opririi posibile a procesului de căutare a

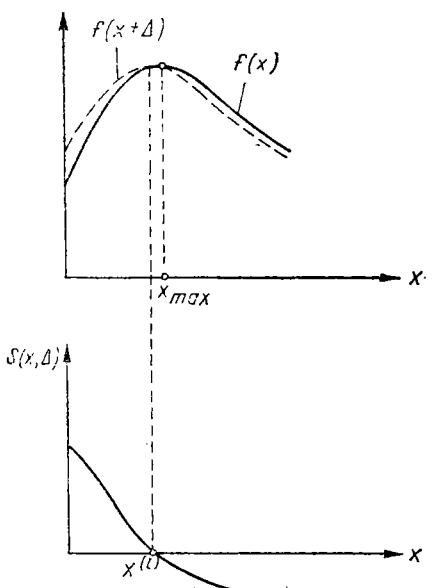


Fig. 5.7

maximului, care este întrucîtva deplasat față de maximul adevărat.

La stînga punctului $x^{(l)}$ înclinarea este pozitivă, iar la dreapta este negativă.

Să considerăm acum formula de bază

$$x^{(l+1)} = x^{(l)} + \theta S\,(x^{(l)},\ \Delta).$$

Acțiunea ei se poate descrie cu ușurință în limbaj geometric. Să construim graficele a două funcții:

$$y(x) = x \text{ și } \varphi(x) = x + \theta S\,(x, \Delta).$$

Primul din aceste grafice este pur și simplu bisectoarea unghiului drept, iar al doilea se îndepărtează de ea cu atît mai mult cu cît mai mare l-am ales pe θ și cu cît mai mare este înclinarea lui $S\ (x, \Delta)$. (În figurile 5.8 *a*, *b*, *c* sînt indicate așezările acestor grafice pentru valorile $\theta = 0, 1$; 0,5; 3 și $S\ (x, \Delta)$ calculat după funcția scop reprezentată în figura 5.7).

Graficele $y(x)$ și $\varphi(x)$ se intersectează într-un punct, acolo unde $y(x) = \varphi(x)$, adică pentru $S(x, \Delta) = 0$, care este punctul posibil de oprire $x^{(l)}$.

Să trecem acum la procesul căutării.

Formula fundamentală poate fi scrisă sub forma

$$x^{(l+1)} = y\,[x^{(l+1)}] = \varphi\,[x^{(l)}].$$

Să fixăm valoarea inițială $x^{(0)}$ și să aflăm $\varphi[x^{(0)}]$. Pe grafic acesta este punctul A_0, dar ordonata sa trebuie să fie egală cu $x^{(l)}$, astfel încît, ducînd o dreaptă orizontală din punctul A_0 pînă la intersecția cu bisectoarea și o dreaptă verticală de la această intersecție pînă la curba $\varphi(x)$, ajungem la punctul A_1. Ordonata acestuia este $\varphi[x^{(l)}]$, adică $x^{(2)}$. Astfel procesul căutării este reprezentat de o linie frîntă care se înscrie între graficele $\varphi(x)$ și $y(x)$. Este evident că punctul $x^{(l)}$ ca atare poate fi atins numai printr-un număr infinit de pași, dar ieșirea la o distanță suficient de mică, practic indiscernabilă, pretinde firește doar un efort limitat.

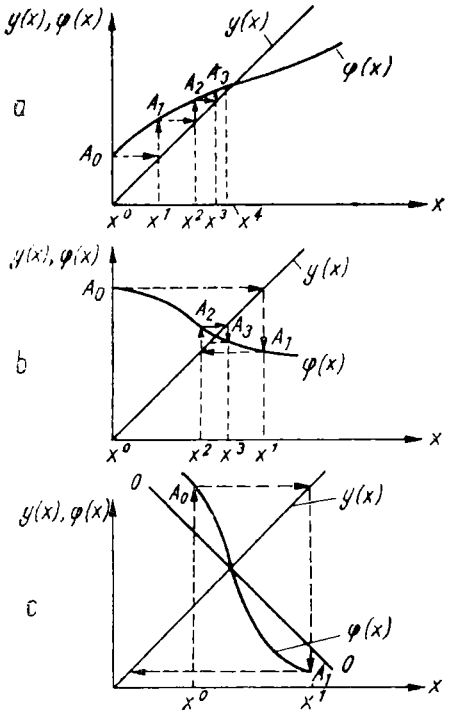


Fig. 5.8

Să trecem la examinarea figurii 5.8 *b*. Aici θ este mai mare, iar curba $\varphi(x)$ este situată la stînga lui $x^{(l)}$, mai sus de nivelul $x^{(l)}$, iar la dreapta, mai jos de acest nivel. Procesul căutării este iarăși reprezentat de o linie frîntă, dar ea nu mai are forma unei scărișoare care se înalță lin pînă la punctul de oprire, ci seamănă cu o pînză de păianjen care se desfășoară în spirală în jurul acestui punct. Punctele opririlor intermediare la fiecare pas se situează rînd pe rînd la dreapta și la stînga. Totuși, procesul de apropiere de maxim decurge aici chiar ceva mai repede, deși, și aici, punctul $x^{(l)}$ este doar limita eforturilor, o limită

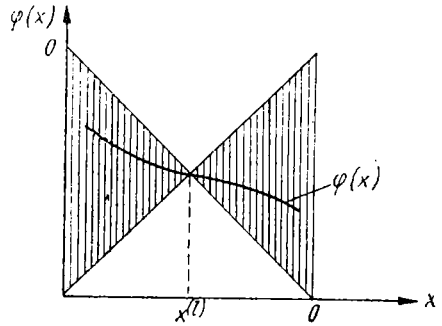


Fig. 5.9

inaccesibilă pentru un număr finit de pași. Dar după cum am mai spus, aceasta nu-i o pagubă prea mare.

Prost de tot stau lucrurile în cazul corespunzător figurii 5.8 *c*, unde $\theta = 3$. Aici curba $\varphi(x)$ s-a desfășurat într-atît încît trece peste dreapta *OO*, care este perpendiculară pe bisectoare (adică face un unghi de — 45° cu axa absciselor și deci are înclinarea egală cu — 1). Așa stînd lucrurile, linia frîntă care reprezintă procesul aproximărilor succesive capătă forma unei spirale desfășurate.

Toate aceste construcții geometrice ne permit să formulăm o concluzie importantă. Dacă graficul funcției $\varphi(x)$ *într-o vecinătate a punctului* $x^{(1)}$ *se află în sectorul dintre bisectoare și perpendiculara pe ea OO (fig.* 5.9, *partea hașurată), atunci procesul căutării permite apropierea nelimitată de acest punct. Dacă însă graficul se află în întregime în exteriorul sectorului indicat, procesul aproximațiilor succesive ne îndepărtează de punctul respectiv*.*

De altfel, acum își găsește ușor explicația și faptul că nu există vreun pericol real de a ajunge, în urma procesului

---

* O expunere mai aprofundată, deși accesibilă pentru elevii care cunosc bine matematica, a problemei convergenței metodei aproximațiilor succesive se poate găsi în prelegerile lui N.I. Vilenkin [25]. (Aici și mai departe numărul dintre parantezele drepte indică numărul de ordine din bibliografia de la sfîrșitul cărții).

de căutare a maximului, în vecinătatea unui minim (situație indicată în fig. 5.3 b).

În apropierea minimului curba $S(x, \Delta)$ trece de asemenea prin zero (fig. 5.10), dar înclinarea este negativă la stînga și pozitivă la dreapta, astfel încît, chiar pentru un θ mic, $\varphi(x)$ va fi mai jos de $y(x)$ la stînga punctului de intersecție și mai sus la dreapta acestui punct, iar procesul căutării în mod obligatoriu ne va îndepărta de el. Pericolul constă numai în a alege ca aproximare inițială tocmai punctul intersecției, ceea ce în cazul unei alegeri arbitrare este practic improbabil.

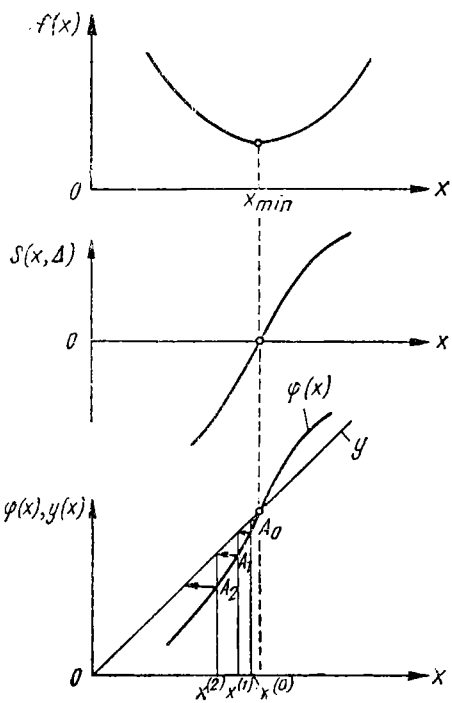


Fig. 5.10

Acum merită să vedem cum va decurge procesul de căutare în cazul cînd funcția-scop are cîteva maxime și minime locale clar exprimate. În figura 5.11 este prezentat graficul unei asemenea funcții și graficul variației înclinării ei, iar în figura 5.12 funcția corespunzătoare $\varphi(x)$ și $y(x) = x$. Domeniul căutării este limitat de abscisele x_{inf} și x_{sup}. Imaginea vorbește de la sine. Fiecare din cele trei maxime locale are propria sa zonă de întindere. Dacă aproximarea inițială este adoptată undeva în domeniul I, atunci procesul de iterare conduce către un punct apropiat de primul maxim, mai exact către un punct unde pentru prima oară înclinarea se anulează. (Această precizare este esențială, întrucît ea amintește de posibilitatea de a sări peste moviliţele mici, în apropierea cărora $S(x, \Delta)$ nu trece prin zero.) Dacă aproximarea inițială aparține zonei II, atunci procesul conduce către cel de-al doilea maxim local. Același lucru se referă și la zona III, dar aici maximul local se află nu în apropierea punctului cu înclinare nulă, ci pe frontiera domeniului admis. Procesul căutării stopează pur și simplu pentru că punctul următor

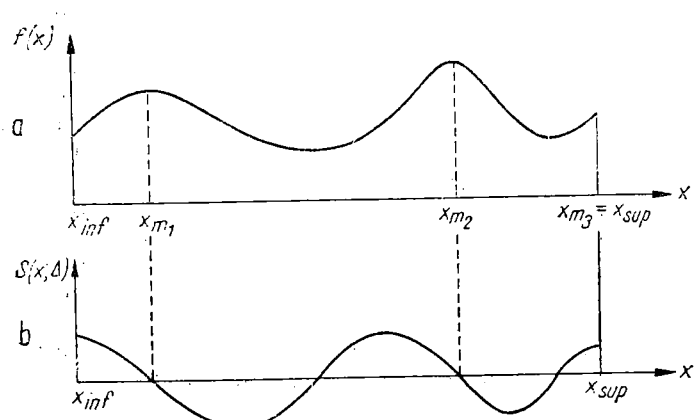


Fig. 5.11

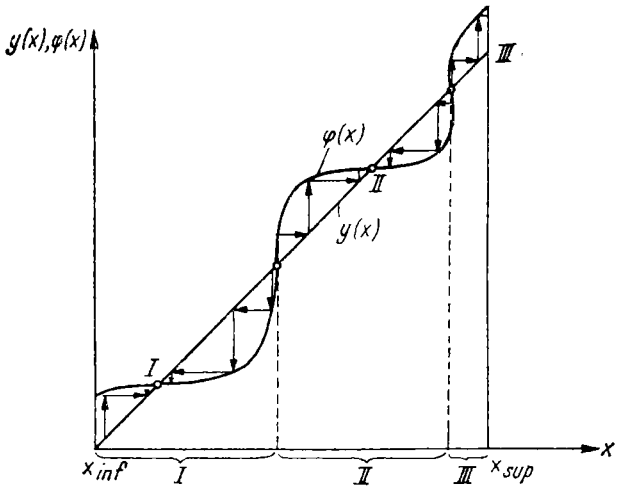


Fig. 5.12

care ar trebui cercetat conform schemei este de fapt în afara accesibilului*.

Granițele care despart zonele de gravitare sînt puncte de minim, mai precis puncte în care înclinarea ce se anulează este negativă la stînga și pozitivă la dreapta.

Firește, figura 5.12 reprezintă o situație cînd coeficientul de proporționalitate θ este ales suficient de bine; dacă însă acest coeficient este foarte mare, situația se poate modifica. În figura 5.13 se arată că în zona II căutarea conduce nu către un punct de maxim, ci către oscilații ciclice în jurul acestui punct la mare distanță de el.

Figurile date sînt foarte intuitive și explică bine întregul proces al căutării. Dar ele nu dau posibilitatea de a se dovedi de ce schema căutării proporționale este mai eficientă decît iterarea simplă cu pas constant. În primul rînd, să observăm că în cazul iterării simple distanța de la

* Particularitățile procesului de căutare în apropierea frontierelor vor fi descrise mai amănunțit în cadrul cap. 9.

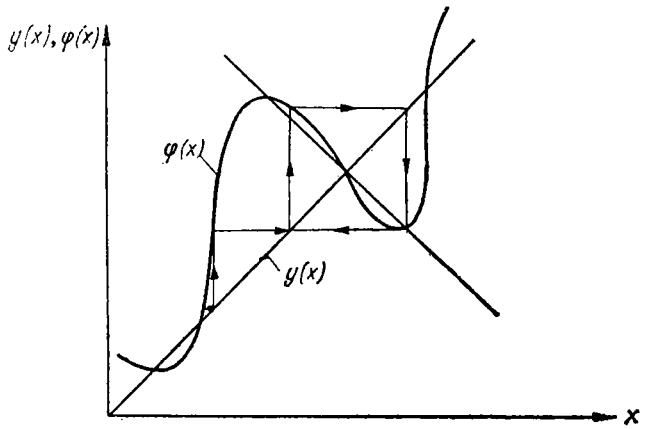


Fig. 5.13

punctul inițial la extremum descrește în progresie aritmetică, adică în n pași descrește cu $n\,\Delta$.

Or, se constată că în cazul căutării proporționale această distanță descrește mai repede decît în progresie geometrică, înmulțindu-se după fiecare pas cu un anumit coeficient mai mic decît rația q fixată.

Dacă acest coeficient este mult mai mic decît unitatea, atunci, odată cu creșterea lui n, reducerea distanței decurge foarte rapid. În figura 5.14 se arată variația distanței la extremum în cadrul utilizării ambelor scheme pe măsura creșterii numărului de iterări ($q = 0{,}5$). Linia plină din această figură corespunde iterării cu pasul constant Δ, iar linia punctată corespunde căutării proporționale. Este limpede că cu cît este mai mare distanța inițială la extremum (în comparație cu Δ) și cu cît q diferă mai mult de unitate, cu atît se manifestă mai viu superioritatea căutării proporționale. Este vorba de o superioritate asemănătoare cu aceea a creșterii dobînzii compuse în comparație cu dobînda simplă!

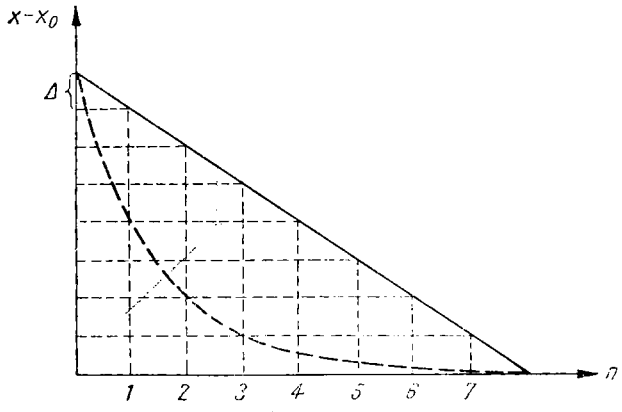


Fig. 5.14

Rămîne de demonstrat că în adevăr căutarea proporțională se supune legilor progresiei geometrice, iar dacă lucrurile stau astfel atunci rămîne de lămurit de cine depinde rația q.

Să revenim la formula de bază

$$x^{(l+1)} = x^{(l)} + \theta\, S(x^{(l)}, \Delta) = \varphi(x^{(l)}).$$

În punctul de oprire $x^{(l)}$ se realizează întotdeauna condiția

$$x^{(l)} = \varphi(x^{(l)}).$$

Scăzînd membru cu membru, obținem:

$$x^{(l+1)} - x^{(l)} = \varphi(x^{(l)}) - \varphi(x^{(l)}).$$

Or, e clar că procesul va fi convergent dacă graficul lui φ nu depășește sectoarele formate de cele două drepte înclinate sub unghiurile de 45° și $-45°$. Să încercăm să-l comprimăm în sectoarele restrînse formate de dreptele cu unghiurile ψ și $-\psi$ (fig. 5.15). Dacă vom reuși, atunci pentru fiece $x^{(l)}$ va avea loc formula

$$\frac{|\varphi(x^{(l)}) - \varphi(x^{(l)})|}{|x^{(l)} - x^{(l)}|} \leqslant \operatorname{tg} \psi.$$

Acum vedem că

$$|x^{(l+1)} - x^{(l)}| = |\varphi(x^{(l)}) - \varphi(x^{(l)})| \leqslant \operatorname{tg} \psi\, |x^{(l)} - x^{(l)}|.$$

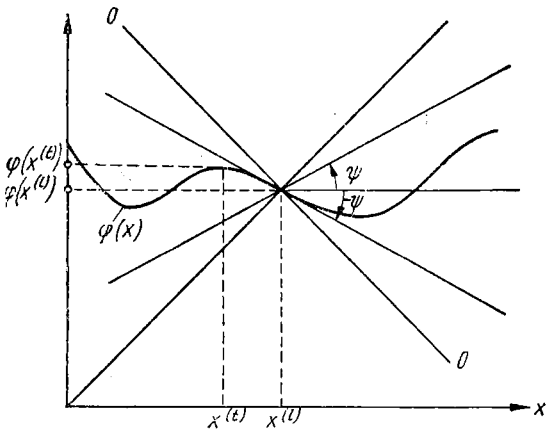


Fig. 5.15

Să notăm cu $\varepsilon^{(t)}$ distanța în cadrul celei de a t-a iterări la extremum, mai precis la oprirea $x^{(l)}$. Atunci $\varepsilon^{(t)}$ este egal cu $|x^{(t)} - x^{(l)}|$ și prin urmare

$$\varepsilon^{(t+1)} \leqslant \operatorname{tg} \psi \cdot \varepsilon^{(t)}$$

De aici rezultă imediat că descreșterea distanței este nu mai înceată decît cea dată de progresia geometrică cu rația q egală cu $\operatorname{tg} \psi$.

Cu cît este mai restrîns sectorul în care reușim să comprimăm funcția $\varphi(x)$, cu atît mai mic este ψ, prin urmare cu atît descreșterea este mai rapidă.

În vecinătatea maximului, graficul înclinării $S(x, \Delta)$ este adesea apropiat de dreapta cu înclinarea $-L$ (fig. 5.16). De aceea funcția $\varphi(x)$ se comportă astfel încît

$$\varphi(x) - \varphi(x^{(l)}) = x - x^{(l)} - \theta L(x - x^{(l)}),$$

de unde rezultă

$$\frac{|\varphi(x) - \varphi(x^{(l)})|}{|x - x^{(l)}|} = |1 - \theta L|,$$

adică în vecinătatea maximului descreșterea distanței la extremum are o viteză nu mai mică decît descreșterea după progresia geometrică cu rația $|1 - \theta L|$.

Cu aceasta am lămurit pe deplin rolul coeficientului θ.

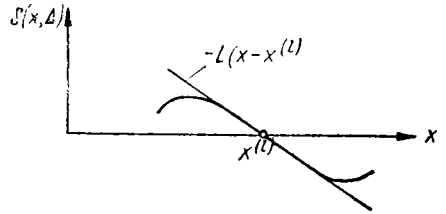


Fig. 5.16

Pentru reușita căutării este necesar ca mărimea lui θ să fie asociată mărimii înclinării curbei de înclinare în apropierea punctului de extremum. Cu cît această înclinare este mai mare, cu atît valoarea coeficientului θ trebuie să fie mai mică, și invers.

Și astfel ajungem la întrebarea: oare n-am putea renunța la alegerea din capul locului a unei valori constante pentru toți pașii a coeficientului θ, oare nu merită să-l modificăm coordonîndu-l cu modificările înclinării înclinării?

Această idee utilă nu este nouă; ea se află la baza metodei de căutare propusă încă de marele Newton.

Să-l alegem pe θ invers proporțional cu înclinarea înclinării, adică cu mărimea

$$w(x, \Delta) = \frac{1}{\Delta}[S(x + \Delta, \Delta) - S(x, \Delta)] =$$

$$= \frac{1}{\Delta}\left[\frac{f(x + 2\Delta) - f(x + \Delta)}{\Delta} - \frac{f(x + \Delta) - f(x)}{\Delta}\right] =$$

$$= \frac{1}{\Delta^2}[f(x + 2\Delta) - 2f(x + \Delta) + f(x)]$$

Mai precis, vom construi aproximări succesive după schema

$$x^{(t+1)} = x^{(t)} + \theta^{(t)} S(x^{(t)}, \Delta),$$

unde

$$\theta(t) = -\frac{\alpha}{w(x^{(t)}, \Delta)}, \quad \alpha > 0.$$

Semnul minus din ultima egalitate se explică prin aceea că în apropierea maximului înclinarea are o pantă negativă, adică w este mai mic decît zero; or, noi trebuie să asigurăm un coeficient θ pozitiv. În ce privește α, acesta poate fi luat pur și simplu egal cu unitatea, după cum recomanda Newton, sau mai mic decît unitatea dacă vrem să fim mai prudenți. E drept că pentru a calcula mărimea $w(x, \Delta)$ va fi nevoie nu de o singură măsurătoare, ci de două măsurători în apropierea fiecărui punct, dar, din păcate, în căutare nici un fel de perfecționare nu se poate obține în mod gratuit.

În schimb, priviți figura 5.17: utilizînd schema lui Newton, căutarea conduce la maxim practic dintr-un singur pas, și aceasta este lesne de înțeles; doar acum punctul $x^{(1)}$ se află după regula ($\alpha = 1$)

$$x^{(1)} = x^{(0)} - \frac{1}{w(x^{(0)}, \Delta)} \cdot S(x^{(0)}, \Delta)$$

de unde rezultă că

$$\frac{S(x^0, \Delta)}{x^{(1)} - x^0} = -w(x^{(0)}, \Delta) = \frac{S(x^{(0)}, \Delta) - S(x^{(0)} + \Delta, \Delta)}{\Delta}$$

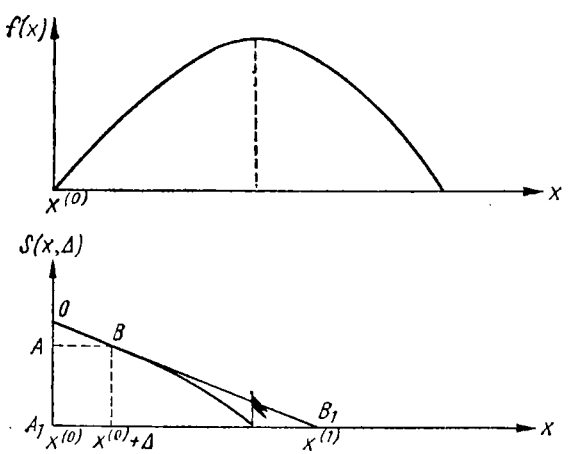


Fig. 5.17

Prin urmare, triunghiurile dreptunghice OAB și OA_1B_1 sînt asemenea și deci punctul $x^{(1)}$ trebuie să se afle la intersecția unei drepte avînd o pantă egală cu înclinarea înclinării w $(x^{(0)}, \Delta)$ și axa absciselor. Dar deoarece înclinarea $S(x, \Delta)$ de la $x^{(0)}$ pînă la maxim este aproape constantă, curba și dreapta aproape că nu se deosebesc, ceea ce și asigură succesul care, în treacăt fie spus, este prezis și de formulele generale pentru evaluarea ritmului de convergență. Firește, așa ceva nu se întîmplă întotdeauna. Din figura 5.18 se vede că, dacă panta $S(x, \Delta)$ se modifică mult, atunci pentru $\Delta = 1$ schema lui Newton ne îndepărtează mult dincolo de punctul dorit. Tocmai de aceea în calcule se preferă de obicei un coeficient de prudență, de exemplu $\alpha = 1/2$. De altfel, ideea de a introduce un coeficient de proporționalitate θ variabil după pas se poate realiza și mai simplu decît în schema lui Newton, unde sînt necesare și măsurători noi și calcule noi.

Într-adevăr, putem începe cu o mărime $\theta^{(0)}$ arbitrară, dar suficient de mare. Dacă rezultatul este fericit, adică pasul conduce la punctul $x^{(1)}$, unde valorile funcției-scop cresc, atunci valoarea coeficientului se menține și în pasul următor; dacă însă ne lovim de insucces, atunci luăm $\theta^{(1)} = \alpha\,\theta^{(0)}$, cu $\alpha < 1$. Micșorînd de fiecare dată pe α în caz de insucces, ne asigurăm împotriva apariției unui exces de „temperament".

Adesea este util să îmbinăm ambele idei: în punctul inițial să facem eforturi suplimentare și să calculăm

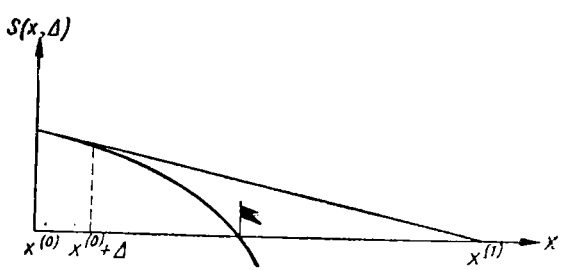


Fig. 5.18

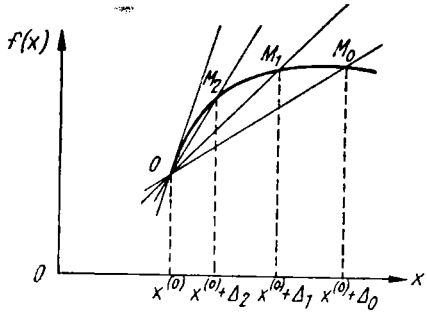


Fig. 5.19

$w(x^{(0)}, \Delta)$, estimînd astfel valoarea rezonabilă pentru $\theta^{(0)}$, iar ulterior să ne călăuzim de procedura simplificată.

Se pare că e timpul să ne oprim: despre procedurile locale de aflare a maximului am vorbit destul de mult. De ce am vorbi numai despre căutarea maximului? **Dar dacă mai binele înseamnă mai puțin?**

Este foarte ușor să răspundem la această întrebare: în toate formulele precedente trebuie doar să înlocuim θ cu $-\theta$ și ele se vor transforma în formule care descriu căutarea minimului*.

Așadar, am putea încheia aici acest capitol, dar autorul simte privirea plină de reproș a cititorului cult (care și-a dat bacalaureatul — secția reală). Cum se poate ca vorbind despre căutarea unui extremum să nu fi pomenit nimic despre ideea de derivată?

Pentru cititorii care încă nu au o pregătire matematică suficientă, este necesară următoarea lămurire. Se numește *derivata* funcției f în punctul x limita către care tinde înclinarea $S(x, \Delta)$ atunci cînd Δ tinde către zero. Să privim figura 5.19. Să măsurăm funcția în punctul $x^{(0)}$ și în punctul $x^{(0)} + \Delta_0$; atunci înclinarea va fi determinată de panta dreptei OM_0. Să luăm acum $\Delta_1 < \Delta_0$ și să estimăm din

* Mai simplu, în loc de a căuta minimul valorilor $f(x)$ trebuie căutat maximul simetricului lui f în raport cu axa absciselor, adică maximul valorilor $-f(x)$.

nou înclinarea; acum ea va fi determinată de panta dreptei OM_1. Pe măsura descreșterii deplasării de la $x^{(0)}$, dreptele a căror pantă determină înclinarea se rotesc tinzînd să ocupe locul limită care revine tangentei la graficul funcției în punctul O, adică pentru $x = x^{(0)}$. Tangenta trigonometrică a unghiului format de tangentă cu axa absciselor este aceea care determină valoarea derivatei în acest punct. Aceasta este, ca să zicem așa, adevărata înclinare locală în punctul $x^{(0)}$, în timp ce pînă acum am avut de-a face numai cu *estimații* ale înclinării, cu valorile ei medii în apropierea acestui punct, valori care depindeau de alegerea pasului Δ.

De obicei, derivata se notează cu

$$\frac{df}{dx}$$

sau cu $f'(x)$. În general vorbind, valorile ei sînt diferite în diferitele puncte x, iar dacă vrem să subliniem că este vorba despre un anume punct, de exemplu $x = x^{(0)}$, atunci scriem sau $\left(\frac{df}{dx}\right)_0$ sau $f'(x^{(0)})$.

Este clar din capul locului că, dacă am ști să aflăm derivata, am putea face uz cu ușurință de formulele căutării înlocuind estimațiile înclinării $S(x, \Delta)$ cu înclinarea reală, adică derivata $f'(x)$.

Nu totdeauna există o atare posibilitate. Dacă problema căutării este de asemenea natură încît valorile funcției-scop pot fi măsurate numai în anumite puncte, nu se poate inventa nimic în afară de estimarea înclinării. Într-adevăr, numai din punct de vedere matematic putem să-l facem pe Δ să tindă către zero; cînd măsurăm, trebuie întotdeauna să luăm pentru Δ o valoare diferită de zero, fie ea oricît de mică. Dar după cum am mai spus, există multe probleme de căutare a extremumului pentru funcții care sînt date cu ajutorul unei formule analitice. Or, dacă avem o formulă care stabilește valoarea funcției-scop pentru orice Δ dat, atunci probabil că vom putea afla și formula corespunzătoare derivatei; în realitate, tocmai astfel stau lucrurile.

Fără a încerca să stăruim asupra tuturor subtilităților procesului de deducere a formulelor pentru derivate, proces care în matematică este numit *derivare*, să prezentăm deocamdată problema cea mai simplă a derivării, derivarea polinoamelor de grad n, adică a funcțiilor date sub forma

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0,$$

unde a_0, a_1, ..., a_n sînt mărimi constante.

Pentru aceasta, să arătăm mai întîi că pentru orice număr întreg k avem egalitatea

$$(x + \Delta)^k = x^k + k\Delta x^{k-1} + \ldots, \quad k = 1, 2, \ldots,$$

unde toți termenii neexplicitați, dacă există asemenea termeni, conțin Δ la o putere mai mare decît unitatea (adică Δ^2, Δ^3,...).

Pentru a demonstra acest lucru, se poate face uz de *metoda inducției matematice*. Să presupunem că formula este adevărată pentru un anumit k, $k = n$. Atunci

$$(x + \Delta)^{n+1} = (x + \Delta)(x + \Delta)^n =$$

$$= (x + \Delta)(x^n + n\Delta x^{n-1} + \ldots) = x^{n+1} + \Delta x^n + n\Delta x^n + \ldots,$$

unde iarăși toți termeni nexplicitați conțin Δ la o putere mai mare decît unitatea. Prin urmare,

$$(x + \Delta)^{n+1} = x^{n+1} + (n + 1)\,\Delta x^n + \ldots,$$

adică: dacă formula este adevărată pentru $k = n$, atunci ea rămîne adevărată și pentru $k = n + 1$. Întrucît pentru $k = 1$ ea este evidentă, rezultă că formula este valabilă pentru orice k.

Acum putem alcătui o expresie și pentru înclinarea funcției dată prin formula $f(x) = x^k$:

$$S(x, \Delta) = \frac{1}{\Delta}\left[(x + \Delta)^k - x^k\right] =$$

$$= \frac{1}{\Delta}\left[-x^k + x^k + \Delta k x^{k-1} + \ldots\right] = kx^{k-1} + \ldots$$

unde ceilalți termeni sînt proporționali cu puterile lui Δ, prin urmare vor tinde către zero odată cu Δ. Să observăm că această formulă rămîne valabilă și atunci cînd Δ va fi o mărime negativă.

În concluzie:

dacă $f(x) = x^k$, atunci $f'(x) = kx^{k-1}$.

Ne putem convinge fără dificultate că, dacă funcția este înmulțită cu un număr constant, atunci și înclinarea ei se va înmulți cu aceeași constantă. Dar aceasta rămîne valabil și pentru derivată, în calitate de limită a înclinării. Prin urmare,

dacă $f(x) = a_k x^k$, atunci $f'(x) = ka_k x^{k-1}$.

Este la fel de evident că derivata unei sume de funcții va fi egală cu suma derivatelor termenilor respectivi. Așadar,

dacă $f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$,

atunci $f'(x) = na_n x^{n-1} + (n-1)a_{n-1} x^{n-2} + \ldots + a_1$.

În particular, derivata constantei este nulă, iar derivata funcției *liniare* dată prin formula $f(x) = a_1 x + a_0$ este egală cu a_1, adică este constantă; lucru firesc, întrucît graficul funcției liniare reprezintă o dreaptă care formează un unghi constant cu axa absciselor (fig. 5.20).

Formulele pentru derivarea altor funcții pot fi aflate cu ajutorul teoriei limitelor. De altfel, le puteți găsi în orice memorator.

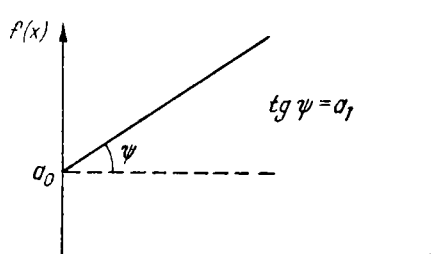


Fig. 5.20

Să vorbim acum mai amănunțit despre *rolul derivatelor în problema căutării extremumului.*

În primul rînd, vom aminti că din însăși definiția înclinării rezultă

$$f(x + \Delta) = f(x) + \Delta \cdot S(x, \Delta).$$

Pe de altă parte, din însăși semnificația noțiunii de derivată, ca valoare limită a înclinării, este clar că

$$f'(x) = S(x, \Delta) + \ldots,$$

unde termenii nescriși tind către zero, atunci cînd Δ tinde către zero. Deci

$$f(x + \Delta) = f(x) + \Delta f'(x) + \ldots$$

Pentru polinoame, această relație este valabilă și pentru $\Delta < 0$. La fel și pentru orice funcții regulate (fără frînger).

Se poate demonstra următoarea importantă afirmație: **dacă în punctul $x = x^{(0)}$ care nu se află la frontiera domeniului căutării se atinge un extremum local al funcției $f(x)$, atunci în acest punct valoarea derivatei, $f'(x^{(0)})$, este egală cu zero.**

Într-adevăr, dacă în punctul $x = x^{(0)}$ se află un maxim, atunci pentru orice punct cu coordonata $x^{(0)} + \Delta$, oricare ar fi Δ, trebuie să fie îndeplinită condiția

$$f(x^{(0)}) \geqslant f(x^{(0)} + \Delta).$$

Dar avem formula

$$f(x^{(0)} + \Delta) = f(x^{(0)}) + \Delta f'(x^{(0)}) + \ldots,$$

unde termenii nescriși devin neglijabili în comparație cu cei scriși atunci cînd Δ descrește.

Înseamnă că, dacă $x^{(0)}$ este coordonata maximului, atunci

$$f(x^{(0)}) \geqslant f(x^{(0)} + \Delta) = f(x^{(0)}) + \Delta f'(x^{(0)}) + \ldots,$$

adică $\Delta f'(x^{(0)}) \leqslant 0$.

Să amintim însă că mărimea Δ, fiind mică, poate primi orice semn (punctul învecinat se poate situa atît la stînga cît și la dreapta lui $x^{(0)}$). Să presupunem că $f'(x^{(0)})$ este un număr pozitiv. Atunci condiția pe care am scris-o va fi încălcată pentru valorile pozitive ale lui Δ. Dacă

însă presupunem că $f'(x^{(0)})$ este un număr negativ, atunci încălcarea va avea loc pentru valorile negative ale lui Δ (la stînga lui $x^{(0)}$). Așadar, rămîne unica posibilitate $f'(x^{(0)}) = 0$, ceea ce dovedește afirmația noastră.

Derivata determină panta tangentei. Prin urmare, informația noastră se reduce la următorul fapt evident din punct de vedere geometric: **într-un punct de maxim tangenta trebuie să fie orizontală(fig. 5.21). Același lucru este adevărat și pentru un punct de minim nesituat pe frontiera domeniului căutării.**

Dacă însă extremumul local se află pe frontieră, atunci valoarea sa se poate compara numai cu valorile punctelor vecine situate de o singură parte (de exemplu pentru un extremum situat pe frontiera stîngă a domeniului de definiție a funcției se iau în considerare numai punctele vecine la dreapta). De aceea, un raționament întru totul analog cu cel de care am făcut uz ne conduce la afirmații de tipul următor: dacă în punctul $x^{(0)}$ situat pe frontiera superioară se obține un maxim, atunci $f'(x^{(0)}) \geqslant 0$. De altfel, atît această afirmație cît și altele referitoare la punctele de frontieră sînt din punct de vedere geometric întru totul clare (a se vedea fig. 5.21).

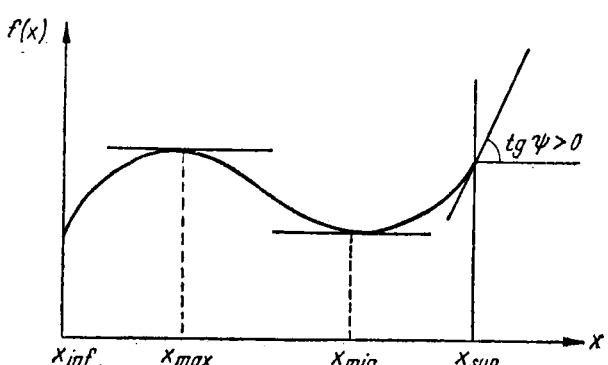


Fig. 5.21

Să revenim acum la studierea procedurii de căutare locală care face uz de cunoașterea derivatei în fiecare punct:

$$x^{(t+1)} = x^{(t)} \pm \theta f'(x^{(t)}), \quad \theta > 0$$

unde se pune semnul plus atunci cînd se caută maximul și semnul minus atunci cînd se caută minimul. În ambele cazuri, procesul poate stopa fie pentru că se lovește de frontieră, fie într-un asemenea punct $x^{(t)}$, unde

$$f'(x^{(t)}) = 0.$$

De aceea toate punctele de acest fel se numesc *staționare*, ceea ce înseamnă că sînt puncte de așezare, liniștire, stare constantă. Am mai văzut însă că într-un maxim sau un minim care nu se află pe frontieră condiția de egalitate cu zero a derivatei este obligatorie. Așadar, **toate maximele și minimele sînt puncte staționare.**

Din păcate, **reciproca nu este adevărată: nu orice punct staționar este punct de maxim sau punct de minim.** Există și alte puncte staționare, denumite *puncte de inflexiune*. Priviți figura 5.22. La stînga punctului x_i

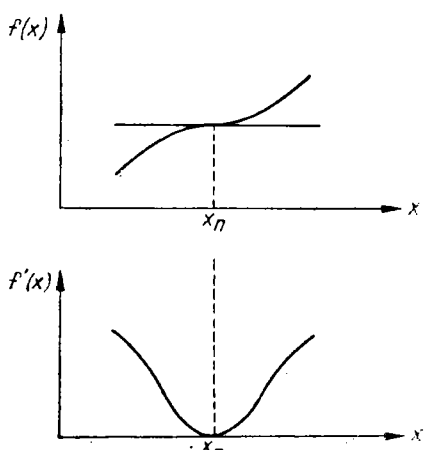


Fig. 5.22

derivata este pozitivă. Apropiindu-ne de el, derivata descreşte ca şi cum ar dori să treacă prin zero şi să devină negativă (în care caz ar apărea un minim), dar, cînd ajungem la zero, derivata nu se hotărăşte să-l treacă şi devine din nou pozitivă. Este clar că punctul de inflexiune este un caz deosebit, degenerat şi nu merită să-i acordăm multă atenţie. Aşadar putem afirma că schema de căutare bazată pe utilizarea derivatelor tinde să stopeze tocmai în extremumurile locale şi în vecinătăţile lor, ceea ce şi explică utilitatea ei*.

Firește, acţiunea schemei depinde de alegerea parametrului θ, la fel ca atunci cînd facem uz nu de derivată, ci de înclinarea în apropierea fiecărui punct. Nu trebuie să mai repetăm cele spuse, dar merită să atragem atenţia asupra unei întrebări care nu e lipsită de interes: mai există oare necesitatea de a construi o schemă de căutare atunci cînd dispunem de formula derivatei?

Într-adevăr, este suficient să rezolvăm ecuaţia $f'(x) = 0$ şi rădăcinile ne vor da coordonatele tuturor punctelor staţionare posibile, prin urmare ale tuturor extremumurilor locale posibile. Întrucît, de regulă, acestea sînt în număr finit şi nu prea mare, valorile funcţiei în toate aceste puncte se pot calcula lesne. Comparînd între ele aceste valori, ajungem de îndată la ţelul final: cea mai mare ne va da maximul absolut, cea mai mică minimul absolut. E drept: această idee este ispititoare la culme şi bine cunoscută, datorită manualelor de analiză pentru liceu, cititorului cult care insistă asupra utilităţii studiului derivatelor.

Din păcate, ideea este minunată, dar... în practică rar se poate realiza integral. Ea se bazează pe ipoteza că putem afla toate rădăcinile ecuaţiei $f'(x) = 0$, iar această ipoteză, din păcate, este prea îndrăzneaţă. Firește, cititorul

---

* Firește, sensibilă cum este la orice extremum local, această schemă care face uz de înclinarea reală pierde însuşirea de a ocoli micile movile şi gropi pe pantă. Să observăm însă şi o împrejurare favorabilă: căutarea maximului nu ne poate conduce la un minim sau la un punct de inflexiune, întrucît în ambele cazuri opririle sînt nestabile faţă de deplasările mici.

a învățat la școală să rezolve ecuațiile algebrice de gradele 1 și 2; probabil că știe și despre existența unor grozave formule pentru rezolvarea ecuațiilor de gradele 3 și 4; dar el știe de asemenea că chiar pentru ecuațiile algebrice de grad mai mare nu se pot stabili, în cazul general, formule explicite pentru calculul rădăcinilor. Aceasta este cu atît mai adevărat pentru ecuații stranii (dar pe care în problemele de căutare le întîlnim efectiv) de felul următor:

$$\cos x - x \cdot 2^x = 0.$$

Firește, de aici nu rezultă că în aceste cazuri nu se pot găsi rădăcinile. Ele pot fi găsite, dar pentru aceasta trebuie *căutate*, făcînd apel la diferite metode de aproximare succesivă.

Merită să ne oprim asupra uneia dintre ele, care pare a fi cea mai simplă. Să presupunem că trebuie să aflăm rădăcinile ecuației

$$g(x) = 0,$$

unde $g(x)$ sînt valori măsurabile sau calculabile ale unei derivate.

Această ecuație se poate transcrie și astfel:

$$\frac{1}{\theta}(x - x) + g(x) = 0, \text{ unde } \theta \neq 0,$$

sau

$$x = x + \theta g(x).$$

O asemenea ecuație este, firește, echivalentă cu cea inițială (de fapt n-am făcut decît să adăugăm zero în membrul stîng), dar pentru scopurile noastre e o formă mai comodă. Să folosim ideea *iterării simple**: să luăm ca aproximare inițială o valoare $x = x^{(0)}$ și, introducînd-o în membrul drept al ecuației, obținem

$$x^{(0)} + \theta g(x^{(0)}).$$

---

* Cititorul care posedă cunoștințe elementare de automatică își dă seama că ideea iterării simple este ideea comenzii cu reacție inversă în raport cu *eroarea g* și avînd un coeficient de amplificare θ.

Dacă această mărime este egală cu $x^{(0)}$, înseamnă că totul este în ordine: $x = x^{(0)}$ satisface ecuația. Dar dacă nu? Atunci o luăm drept a doua aproximare și, punînd

$$x^{(1)} = x^{(0)} + \theta g(x^{(0)}),$$

verificăm dacă ecuația nu va fi satisfăcută pentru $x = x^{(1)}$.

Continuăm această procedură de iterare, acționînd de fiecare dată după regula: dacă aproximarea curentă este $x^{(t)}$, atunci următoarea se calculează după formula

$$x^{(t+1)} = x^{(t)} + \theta g(x^{(t)}).$$

Dacă se va constata că pentru o anume iterare $t = l$ avem $x^{(l+1)} = x^{(l)}$, atunci, firește, procesul va stopa; am găsit valoarea rădăcinii. Rolul mărimii θ este clar: cu cît θ este mai mic, cu atît căutarea se face în pași mai mici, adică mai prudent. Este evident că pasul depinde și de valoarea funcției în punctul respectiv, lucru rațional: cu cît abaterea de la zero este mai mare, cu atît eroarea este mai grosolană și punctul unde ea a avut loc trebuie părăsit cît mai repede.

Dar procedura iterării simple poate fi considerată și din alt punct de vedere. Dacă g reprezintă derivata funcției f, adică $g(x) = f'(x)$, atunci formula iterării simple se poate transcrie în felul următor:

$$x^{(t+1)} = x^{(t)} + \theta f'(x^{(t)}).$$

Or, aceasta este chiar formula cunoscută din metoda proporțională a căutării extremumului funcției!

Așadar, am stabilit că metoda iterării simple în cazul căutării rădăcinii ecuației

$$f'(x) = 0$$

este echivalentă cu metoda proporțională a căutării extremumului funcției f. Este cunoscută și altă metodă de căutare a rădăcinii, în cadrul căreia la fiecare pas t se calcu-

leză valoarea derivatei funcției g, adică $g'(x^{(t)})$ și se utilizează nu un coeficient de proporționalitate constant, ci

$$\theta^{(t)} = -\frac{1}{g'(x^{(t)})}.$$

Dar întrucît g este ea însăși derivata (înclinarea reală) a unei funcții f, g' este derivata acestei derivate, adică înclinarea înclinării. Prin urmare, această metodă coincide cu metoda lui Newton, pe care o cunoaștem deja — desigur, dacă în formula de acolo înlocuim $w\ (x, \Delta)$ cu derivata derivatei, care este denumită derivata a doua și care se notează cu f'', sau $\frac{d^2f}{dx^2}$, adică dacă se consideră că

$$x^{(t+1)} = x^{(t)} - \frac{f'(x^{(t)})}{f''(x^{(t)})}.$$

Trebuie însă să recunoaștem deschis că metoda a fost propusă de către Newton tocmai pentru căutarea rădăcinilor și abia mai tîrziu transpusă pentru problemele de căutare a unui extremum.

Așadar, am ajuns la concluzia că, dacă rădăcinile ecuației nu pot fi aflate explicit (și de obicei tocmai așa se întîmplă), trebuie să facem apel la proceduri de căutare a rădăcinilor care de fapt coincid cu procedurile de căutare locală a optimului pe care le cunoaștem*. Dealtfel, aceasta confirmă doar încă o dată forța și caracterul rațional al acestor proceduri, așa că e timpul să le oferim un cîmp mai larg de activitate.

---

* Firește că aici căutarea rădăcinilor are de asemenea un caracter local; după cum alegem o valoare inițială sau alta, ajungem la o rădăcină sau alta. Găsirea tuturor rădăcinilor este o problemă tot atît de dificilă ca și detectarea tuturor extremelor locale.

6
Drumuri în spațiu

(Panta cea mai rapidă)

Cutezătoare gînduri și viziuni
Pe ei îi duc spre zări îndepărtate,
Ei urcă pe cărări pieptișe
Pe care nu se-ncumetă alți muritori.
E. Mejelaitis, *La monumentul Geniilor*

S-a ivit, așadar, o nouă idee, ideea prognozei, adică a extrapolării variației valorilor funcției pe baza rezultatelor studierii comportării funcției într-o vecinătate mică. Să încercăm să facem uz de această metodă pentru rezolvarea „tenebroaselor" probleme multidimensionale.

Cu „doi" a și început „multul"; ne vom limita la cazul bidimensional, unde totul este suficient de intuitiv: domeniul căutării este o parte a planului și se poate lesne figura pe hîrtie, iar funcția-scop se poate reprezenta întotdeauna sub forma unei movilițe (dacă se caută maximul) sau a unei căldări (dacă se caută minimul)

Să observăm că, deși desenul în spațiu este intuitiv (fig. 6.1), va fi totuși mai comod să folosim figurarea în plan, făcînd apel la procedeul, binecunoscut din hărțile geografice, de reprezentare a altitudinilor cu ajutorul *izohipselor* sau *curbelor de nivel* (fig. 6.2). Pentru fiecare din aceste linii vom nota valoarea funcției (altitudinea!) care este identică pentru toate punctele situate pe linia respectivă *. Iar pentru ca totul să fie asemănător cu o

* De fapt am și făcut uz de o reprezentare cu ajutorul orizontalelor în figura 4.4, unde se înfățișa funcția cu valorile ei pe rețea.

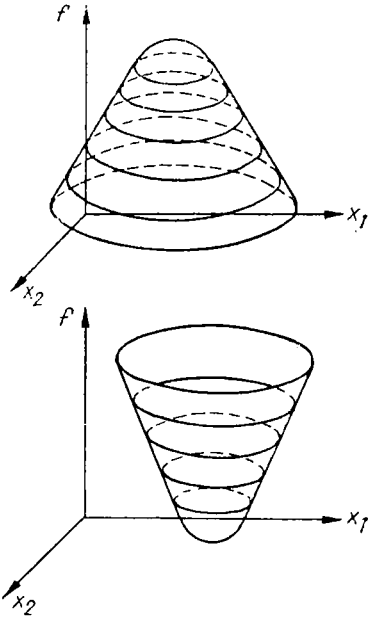


Fig. 6.1

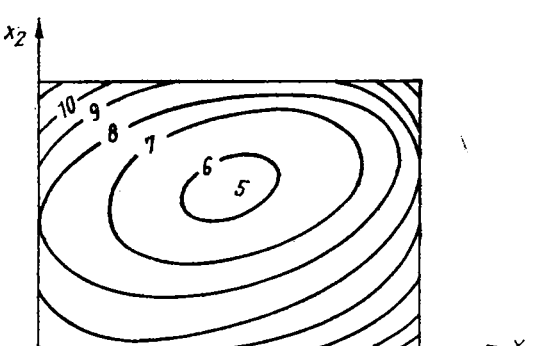


Fig. 6.2

hartă, vom considera că axa x_1 este orientată dinspre apus înspre răsărit, iar axa x_2 dinspre sud înspre nord.

Acum e timpul să pornim: e necesar ca, plecînd din unghiul nord-vestic al depresiunii, să ajungem la fundul ei. Este plictisitor să umbli de unul singur. De aceea vom organiza o competiție sportivă. Vor participa trei maeștri cunoscuți: cu numărul de ordine 1 apare fostul campion mondial Păsescu, zis și Rețeanu, cu numărul 2, prietenul său Prognozea, supranumit Negrăbilă, iar numărul 3 se afla pe pieptul campionului de anul trecut, Cercetașu, poreclit Sarebine. Tactica lor este bine cunoscută spectatorilor care s-au adunat la marginea depresiunii, acoperită cu o iarbă deasă și înaltă. Spectatorii își salută amical favoriții. Dar... mulțimea suporterilor este oarecum intrigată: a apărut și un al patrulea sportiv, pe care nu-l cunoaște nimeni. El seamănă întrucîtva — e greu de spus prin ce anume — cu campionii căutării, și anume seamănă cu toți, dar cine anume este nu se poate ști, deoarece fața îi este ascunsă de o mască neagră. Spectatorii sînt intrigați și așteaptă cu nerăbdare startul. Iată că sportivii s-au așezat în front (ceva mai jos de marginea depresiunii). Arbitrul se apropie de ei și le examinează cu atenție echipamentul: fiecare are o busolă, un altimetru și nimic mai mult. Nici un fel de hărți: hartă are doar arbitrul. Pe ea va marca drumul către finiș al fiecăruia dintre participanți. Dar iată că ultimele pregătiri s-au încheiat. Atenție! Pornirea! Sportivii se cufundă în iarba înaltă, mai înaltă decît omul. În față nu se vede nimic, spectatorii nu-i mai văd pe sportivi, dar se pot observa foarte clar cărările pe care ei și le croiesc.

La început cărările sînt apropiate, apoi se despart. Și, în sfîrșit, vedem că una din ele debușează în petecul de pămînt descoperit al finișului și din iarba înaltă iese în fugă pe poiană un sportiv. Oare n-o fi chiar el, novicele cu masca neagră? Ba da, chiar el este, nimeni nu se mai îndoiește și întreaga vale răsună de aclamațiile mulțimii care-l salută pe noul campion!

Suporterii inițiați l-au și înconjurat pe bărbatul înalt și cărunt care tot timpul s-a ținut deoparte; cu toții în-

cearcă să-i strîngă mîna. Să fie antrenorul Măștii-negre? Nu-i vedem fața, dar la radio se anunță pe un ton solemn: „La competițiile pentru premiul ziarului «Căutarea» a învins tînărul sportiv Mască-neagră! L-am invitat la microfon". După o pauză, se aude: „Vă felicităm pentru marele succes obținut". „Mulțumesc" — răspunde vocea. „Acum puteți să vă dezvăluiți numele adevărat?" „Desigur, mă cheamă Panta-cea-mai-rapidă". „Minunat nume pentru un sportiv". „Firește, îmi place și mie! Dar numele contează mai puțin; toate succesele le datorez antrenorului meu". De fapt, nu merită să ascultăm mai departe. Interviul, prea tradițional, era plicticos; dar numele antrenorului m-a interesat: este un renumit matematician și un sportiv cu stagiu (lucru — în treacăt fie spus — destul de obișnuit pentru matematicienii de azi).

Ce părere aveți, stimate cititor, n-ar fi cazul să ne autoinvităm acasă la un om atît de interesant? Poate că ne va revela secretele tacticii care a contribuit la victoria tînărului său elev...

Pe scurt, ne aflăm în cabinetul profesorului-antrenor și, pentru a nu răpi prea mult timp unei persoane atît de ocupate cu știința și cu sportul, trecem dintr-o dată la chestiune.

Noi: Succesul tînărului dv. elev se datorează probabil utilizării unui nou procedeu tactic?

Antr.: Da, desigur! Dar cu un mic amendament: procedeul este nou în sport, dar în matematică și fizică e folosit de zeci de ani. Voi explica cu plăcere în ce constă — firește, în ipoteza că sînteți familiarizați cu vechiul procedeu.

Noi: Ne este bine cunoscută tactica excampionului Păsescu-Rețeanu, dar pe Prognozea-Negrăbilă și pe Cercetașu-Sarebine i-am văzut numai pe distanțe unidimensionale.

Antr.: Pentru început, aceasta este suficient. De altfel, am păstrat harta pe care sînt marcate etapele traseelor recordmanului și ale celorlalți concurenți. Iat-o în fața dv. (fig. 6.3).

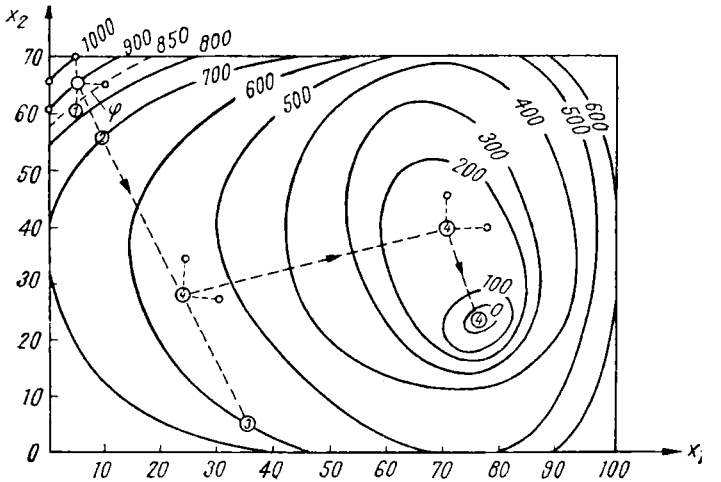


Fig. 6.3

Sportivii și-au luat startul din punctul $x_1^{(0)} = 5$, $x_2^{(0)} = 65$, estimînd altitudinea inițială la $f^{(0)} = 900$ m. Păsescu-Rețeanu a determinat altitudinea în patru puncte, deplasate după axele de coordonate cu 5 m; ceilalți doi s-au limitat la două măsurători vecine*. Coordonatele punctelor $(x_1^{(0)} \pm \Delta;\ x_2^{(0)} \pm \Delta)$ le-am figurat pe hartă, iar apoi am întocmit un tabel special. Vi-l prezint:

x_1	x_2	f
5	65	900
0	65	1000
10	65	850
5	60	840
5	70	1000

* În cele ce urmează cititorul se va putea convinge că și pentru metodele de căutare a căror „personificare" sînt ceilalți sportivi măsurătorile suplimentare pot fi utile.

După executarea măsurătorilor, toți s-au cufundat în calcularea estimării poziției variantei următoare, bănuită a fi un extremum.

Sportivul nr. 1 (Păsescu-Rețeanu) a bănuit dintr-o dată punctul (5; 60) ca fiind cel mai bun din cele cunoscute și a început să examineze stăruitor vecinătatea acestui punct. Cunoașteți bine maniera lui, așa că nu voi insista asupra descrierii ei. Mai ales că nefericitul n-a avut noroc. Vă amintiți probabil că a sosit ultimul?

Noi: Da. Ba chiar circulă zvonul că după această zdrobitoare înfrîngere s-a hotărît să renunțe cu totul la sportul căutării și să se reprofileze pentru loto.

Antr.: Maniera lui s-a cam învechit și mă tem că în orice sport în care are vreun rol prognoza succesele sale nu vor fi mai bune. Dar să revenim la sportivii noștri. Al doilea, Prognozea-Negrăbilă și al treilea, Cercetașu-Sarebine, au acționat într-un mod asemănător cu ceea ce ați avut prilejul de a vedea în căutările unidimensionale. Îmi închipui că vă pricepeți în formulele proceselor de iterare?

Noi: (tăcem cu modestie).

Antr.: Ei bine, ei au folosit aceeași formulă ca în cazul unidimensional, numai că aici a fost necesar să facă uz de ea pentru fiecare coordonată în parte:

$$x_1^{(1)} = x_1^{(0)} - \theta \frac{1}{\Delta} [f(x_1^{(0)} +$$

$$+ \Delta, x_2^{(0)}) - f(x_1^{(0)}, x_2^{(0)})],$$

$$x_2^{(1)} = x_2^{(0)} - \theta \frac{1}{\Delta} [f(x_1^{(0)}, x_2^{(0)} +$$

$$+ \Delta) - f(x_1^{(0)}, x_2^{(0)})].$$

După cum vedeți, mutarea după fiecare coordonată o determină înclinarea, care este estimată prin „pașii de recunoaștere" după aceeași coordonată.

Negrăbilă (permiteți-mi să scurtez întrucîtva bombasticul său nume de familie), considerînd $\theta = 0{,}5$, a obținut:

$$x_1^{(1)} = 5 - \frac{0{,}5}{5}[850 - 900] = 10,$$

$$x_2^{(1)} = 65 - \frac{0{,}5}{5}[1000 - 900] = 55.$$

Sarebine însă, luînd inițial obișnuitul $\theta = 5$, și-a dat repede seama că punctul astfel obținut va fi în afara frontierelor demarcate pentru competiție și, punîndu-și oarecum frînă temperamentului, a trecut la $\theta = 3$. Ca rezultat, a obținut:

$$x_1^{(1)} = 5 - \frac{3}{5}[850 - 900] = 35,$$

$$x_2^{(1)} = 65 - \frac{3}{5}[1000 - 900] = 5.$$

Desigur, ați observat că sportivii nr. 2 și nr. 3, deși s-au oprit în puncte diferite, către obiectivul trasat ei au trebuit să alerge pe aceeași direcție (fig. 6.3). În adevăr, direcția este determinată numai de raportul creșterilor după fiecare dintre coordonate: în ambele cazuri înaintarea a avut loc sub unul și același unghi φ în raport cu axa x_1 (apus — răsărit), și anume

$$\operatorname{tg}\varphi = \frac{f(x_1^{(0)}, x_2^{(0)} + \Delta) - f(x_1^{(0)}, x_2^{(0)})}{f(x_1^{(0)} + \Delta, x_2^{(0)}) - f(x_1^{(0)}, x_2^{(0)})} =$$

$$= -\frac{100}{-50} = 2.$$

Pe același drum a pornit și campionul nostru. Numai că, spre deosebire de ceilalți concurenți, el nu și-a fixat dinainte „gradul de nerăbdare" și a hotărît să-și potrivească nerăbdarea cu împrejurările. Pe drum el examina altimetrul (ceea ce, e drept, i-a consumat o parte din timp) și a alergat numai pînă în momentul cînd a încetat scăderea înălțimii.

Acțiunea sa poate fi relatată și altfel. De fapt, pe drumul său el a testat toate valorile posibile ale lungimii pasului, începînd cu cea mai mică și mărind-o treptat. În felul acesta, pe de o parte, alegînd direcția rațională, campionul a utilizat ideea prognozei, iar pe de altă parte, înaintînd pe direcția aleasă, a realizat pur și simplu trierea. Mișcarea s-a oprit în punctul cel mai avantajos (cel mai coborît pe calea aleasă).

Ei bine, dar ce s-a întîmplat mai departe? Doar știți că fiecare sportiv se ține de principiul iterării, adică după fiecare stopare își repetă procedura: „palpează" versantul din apropiere (fiecare punct de oprire are propriul său versant!), află direcția rațională și înaintează pe ea făcînd uz de principiul său. Asta-i tot! După cum vedeți, destul de simplu!

Noi: Într-adevăr, este simplu. Desigur, e drept că Panta-cea-mai-rapidă a îmbinat în acțiunea sa vechiul (trierea) cu noul (prognoza direcției). Dar, spuneți-mi, totdeauna iese învingător în competiții?

Antr.: Firește că nu. Există și eșecuri. Multe depind de profilul versantului. Știți că atît Negrăbilă cît și Sarebine ghicesc și fixează odată pentru totdeauna ritmul mișcării. Se poate întîmpla ca ritmul dat să corespundă realmente bine întregului profil al versantului. Ei bine, atunci ei învinge, căci nu sînt nevoiți să consume timp pentru a fixa ritmul în cursul înaintării. Dar

poate că acesta e și sensul întrecerii: dacă învingătorul este cunoscut dinainte, de ce să mai organizăm competiții? Pe cînd așa, sportul este sport!

Această frază marca, fără îndoială, încheierea convorbirii, așa că, mulțumind profesorului pentru interviul său, care se transformase într-o prelegere, am plecat acasă. La plecare, profesorul ne-a mai spus cîteva cuvinte care ne-au pus pe gînduri. Iată cuvintele lui: „Aș fi un prost antrenor dacă v-aș dezvălui toate secretele tactice. Panta-cea-mai-rapidă are destule asemenea secrete: el nu este deloc atît de monoton cum poate că vi s-a părut. Dar aș fi totodată și un rău profesor dacă v-aș spune totul, nelăsînd loc pentru gîndire".

Vom mai avea prilejul să reflectăm la aceste probleme pe măsura creșterii cunoștințelor noastre. Dar de pe acum putem să ghicim cîte ceva.

E drept că **ideea celei mai rapide pante constă în aceea ca, „palpînd" vecinătatea și alegînd direcția rațională, să înaintezi exclusiv după această direcție. Această restricție transformă problema căutării multidimensionale într-una unidimensională: rămîne doar să cauți un extremum înaintînd pe dreaptă!** *Or, modalități de căutare unidimensională sînt o mulțime: de la simpla triere pînă la cele mai subtile,* cu care urmează abia să facem cunoștință. Tocmai aici se află rezervele tactice la care făcea aluzie onoratul profesor-antrenor și datorită cărora favoritul său Panta-cea-mai-rapidă își nuanțează comportarea!

S-ar părea, așadar, că totul este clar. Dar cititorul atent are replica pregătită: „Parcă ați vorbit despre probleme multidimensionale, dar de fapt nu s-au discutat decît cele bidimensionale. Ei bine, nu ne luați chiar așa, doar nu sîntem niște primitivi!"

Mult stimatul cititor serios are și nu are dreptate. Căutarea bidimensională am ales-o numai din motive de intuitivitate: nimeni n-ar putea să-și închipuie că vizionează o alergare, o cursă în spațiul cu cinci dimensiuni, dar principiul rămîne același oricare ar fi numărul de dimensiuni. Firește, aici nu mai putem evita formulele.

Să presupunem că are loc o căutare în spațiul n-dimensional (unde n este un număr întreg pozitiv oarecare). Ca întotdeauna, să începem cu un punct ales după voie A_0. Poziția acestui punct în spațiu este determinată de n numere, coordonatele $x_1^{(0)}$, $x_2^{(0)}$, ..., $x_n^{(0)}$. Nu trebuie de loc să încercăm a ni le reprezenta ca în geometrie. Fiecare coordonată poate să aibă sensul său propriu, de exemplu x_1 poate să fie temperatura aerului injectat în cuptor, x_2 — consumul său în volum, x_3 — umiditatea etc. Într-un asemenea caz, coordonatele punctului inițial vor putea fi, de exemplu:

$$x_1^{(0)} = 1000°;\ x_2^{(0)} = 10\ \text{m}^3/\text{min};\ x_3^{(0)} = 40\%.$$

Aici n-are ce căuta geometria obișnuită cu măsurile ei de lungime. Este drept că cel mai bine este să lucrăm cu mărimi relative, *adimensionale*. Dacă marginea inferioară a temperaturilor pentru care este indicată căutarea este $x_{1\,inf} = 500°$, iar marginea superioară este $x_{1\,sup} = 1500°$, atunci este comod să introducem temperatura relativă, definită prin formula

$$\bar{x}_1 = \frac{x_1 - x_{1\,inf}}{x_{1\,sup} - x_{1\,inf}}.$$

Este evident că această temperatură relativă poate varia în punctele domeniului de căutare numai între 0 și 1*. În punctul inițial $x_1^{(0)} = 1000°$, iar valoarea nouă, relativă va fi:

$$\frac{1000° - 500°}{1500° - 500°} = 0{,}5.$$

---

* Firește, cititorul știe că orice temperatură, oricum ar fi definită, este relativă. Cînd măsurăm temperatura în grade Celsius, o calculăm în procente raportate la distanța de la punctul de îngheț al apei pînă la punctul ei de fierbere. Însă pentru a indica temperatura multor fenomene trebuie să „ieșim" din cadrul fixat, considerînd atît temperaturi negative cît și temperaturi peste 100°C. Sistemul pe care-l propunem aici se reduce pur și simplu la faptul că pentru fiecare problemă se face uz de un sistem adecvat de unități, de o „temperatură" adecvată, care variază numai de la 0 pînă la 1.

Tot astfel este convenabil să recalculăm și celelalte coordonate utilizînd pentru coordonata a i-a formulă *

$$\bar{x}_i = \frac{x_i - x_{i\,inf}}{x_{i\,sup} - x_{i\,inf}}.$$

Atunci toate mărimile relative vor varia numai între 0 și 1, adică domeniul căutării devine cubul n-dimensional cu latura unitate.

În ceea ce urmează vom considera că o asemenea recalculare s-a efectuat de la bun început și nu vom mai bara literele x_i pentru a nu îngreuia formulele.

Să revenim acum la problema căutării, de astă dată călăuzindu-ne de ideile „bidimensionale" pe care le-am asimilat. În primul rînd, trebuie să realizăm „palparea" vecinătății punctului inițial. Facem aceasta mărind și micșorînd pe rînd fiecare coordonată cu o mărime mică Δ_i, de exemplu $\Delta_i = 0{,}01$ și măsurînd efectul respectiv (în exemplul considerat efectul ar fi productivitatea cuptorului, măsurată în tone/per oră).

Să revenim acum la limbajul abstract. Așadar, „palparea" vecinătății a constat în măsurarea valorilor funcției în punctele învecinate, ceea ce permite aflarea înclinării după fiecare coordonată:

$$\frac{1}{\Delta_1}[f(x_1^{(0)} + \Delta_1, x_2^{(0)}, \ldots, x_n^{(0)}) - f(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})]$$

$$\frac{1}{\Delta_2}[f(x_1^{(0)}, x_2^{(0)} + \Delta_2, \ldots, x_n^{(0)}) - f(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})],$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\frac{1}{\Delta_n}[f(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} + \Delta_n) - f(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})].$$

Îndată ce înclinările au fost găsite, ne putem deplasa într-un punct nou, efectuînd, după fiecare coordonată, cîte un pas „mare". Mărimea și direcția principală a pasului

* Formula cu numărul de ordine i (*N.T.*).

sînt determinate de mărimea și semnul înclinării, precum și de coeficientul de proporționalitate θ, unic pentru toate coordonatele.

Odată ales acest coeficient, coordonatele noului punct, ale noului regim, devin:

$$x_1^{(1)} = x_1^{(0)} + \theta \cdot \frac{1}{\Delta_1}[f(x_1^{(0)} +$$

$$+ \Delta_1, x_2^{(0)}, \ldots, x_n^{(0)}) - f(x_1^{(0)}, \ldots, x_n^{(0)})]$$

$$x_2^{(1)} = x_2^{(0)} + \theta \cdot \frac{1}{\Delta_2}[f(x_1^{(0)}, x_2^{(0)} +$$

$$+ \Delta_2, \ldots, x_n^{(0)}) - f(x_1^{(0)}, \ldots, x_n^{(0)})]$$

$$x_n^{(1)} = x_n^{(0)} + \theta \cdot \frac{1}{\Delta_n}[f(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} +$$

$$+ \Delta_n) - f(x_1^{(0)}, \ldots, x_n^{(0)})].$$

E drept că toate aceste formule sînt întru totul asemănătoare, oricîte coordonate ar fi: doar două sau peste o mie (tocmai de aceea și spunem că doi este „mult"!).

Dar cum să alegem coeficientul de proporționalitate θ? Cunoaștem deja anumite procedee: ori θ este ales în prealabil o dată pentru totdeauna, pentru întreaga procedură de iterare, pentru orice pas; ori la fiecare pas se alege valoarea cea mai bună $\theta = \theta(t)$, pe calea căutării *unidimensionale* (căci θ nu este decît un număr, un coeficient) măsurînd valorile funcției atunci cînd θ crește începînd de la zero, adică cînd ne mișcăm după direcția aleasă pornind din punctul inițial, și alegînd acel θ care corespunde celei mai bune dintre valorile întîlnite ale funcției. De fapt, pentru aceasta nu este necesar să cunoaștem exact direcția, adică traseul punctului în spațiul n-dimensional atunci cînd θ variază. Putem pur și simplu să determinăm după formule

cum se schimbă fiecare coordonată atunci cînd se schimbă θ*:

$$x_1(\theta) = x_1^{(0)} + \theta \cdot \frac{1}{\Delta_1}[f(x_1^{(0)} + \\ + \Delta_1, x_2^{(0)}, \ldots, x_n^{(0)}) - f(x_1^{(0)}, \ldots, x_n^{(0)})]$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n(\theta) = x_n^{(0)} + \theta \cdot \frac{1}{\Delta_n}[f(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} + \\ + \Delta_n) - f(x_1^{(0)}, \ldots, x_n^{(0)})],$$

și avînd valorile coordonatelor să măsurăm (sau să calculăm) valorile funcției-scop. Prin aceasta, fiecărui θ i se pune în corespondență o anumită valoare a funcției-scop. Așadar, cîtă vreme coordonatele variază nu arbitrar, ci în conformitate riguroasă cu formulele scrise, funcția-scop apare ca o funcție de un singur argument θ, care joacă astfel rolul de coordonată ce măsoară mișcarea pe drumul ales prin tatonare.

Iată de ce problema căutării celui mai bun θ *este realmente unidimensională.*

Sper că cititorul curios (și tocmai unui asemenea cititor i se adresează lucrarea) va dori totuși să afle *ce fel de drum este cel pe care îl străbate punctul în spațiul multidimensional pe măsura creșterii lui* θ. Ca să nu ne rătăcim în labirintul geometriei multidimensionale, vom răspunde că *drumul este pur și simplu o dreaptă, o rază ce pornește din punctul inițial, după cum am și putut vedea pe viu atunci cînd „mult" era doar „doi"*. Dar, pentru a ridica nivelul de cultură, să mai introducem unele noțiuni matematice. Cînd, sondînd vecinătatea, determinam o anumită mărime finită Δ_i, noi de fapt calculam înclinarea medie după fiecare coordonată. Dacă vom face ca acest Δ_i să tindă către zero, vom obține înclinarea în direcția fie-

* În aceste formule totul este determinat de rezultatele cercetării, ale „palpării", cu excepția lui θ, singurul care variază.

cărei coordonate exact în punctul inițial. Un asemenea procedeu l-am mai folosit și în cazul unidimensional, unde vorbisem despre derivată și-o notasem cu $\frac{df}{dx}$. În cazul multidimensional, înclinările după fiecare coordonată într-un punct dat se numesc *derivate parțiale* și se notează $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f^*}{\partial x_n}$; pentru a sublinia în ce anume punct s-a calculat derivata, se pune, de exemplu, $\left(\frac{\partial f}{\partial x_i}\right)_0$. Dacă în formulele pentru determinarea variației coordonatelor pe măsura creșterii lui θ se face uz nu de înclinările medii, ci de derivatele parțiale calculate în punctul inițial, atunci firește că vom înainta după raza ce pornește din acest punct. Direcția acestei raze se numește direcția *gradientului* (în limba latină *gradior, gradii, gressus sum* înseamnă a păși, a merge), ceea ce, în treacăt fie spus, are aproximativ același sens ca și cuvîntul „înclinare". Matematicienii au demonstrat că direcția gradientului este într-un anume sens cea mai bună direcție atunci cînd se caută un maxim**.

Pentru a înțelege cum stau lucrurile, să revenim deocamdată la cazul bidimensional, intuitiv. Să reprezentăm „topografia" funcției-scop cu ajutorul liniilor de nivel (fig. 6.4).

Dacă vom considera mica porțiune din acest desen demarcată de dreptunghiul punctat și o vom reprezenta pe o scară mai mare și cu o desime mai mare a liniilor

---

* Potrivit definiției, formulele se pot scrie (pentru $i = 1, 2, \ldots n$):

$$\frac{\partial f}{\partial x_i} = \lim_{\Delta \to 0} \frac{1}{\Delta} [f(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)].$$

Derivatele parțiale se calculează, evident, la fel ca derivata obișnuită, unidimensională, numai că în calcul trebuie avut în vedere că celelalte coordonate, rămînînd neschimbate, pot fi considerate drept mărimi constante.

** Cînd se caută minimul, atunci dimpotrivă trebuie să înaintăm în direcția inversă gradientului (direcția antigradientului) introducînd în formulă în locul semnului + semnul —.

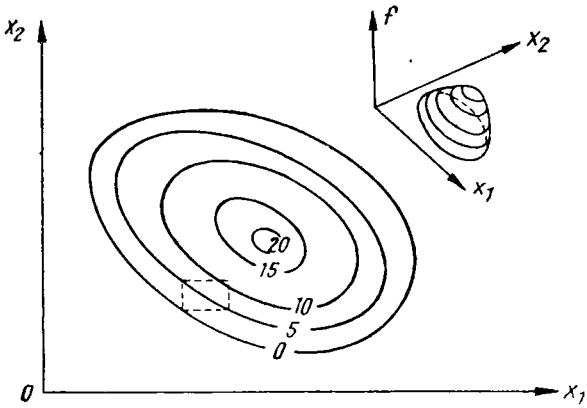
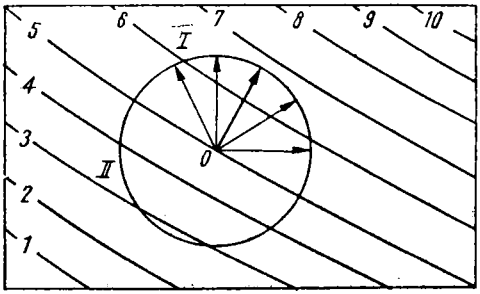


Fig. 6.4

(fig. 6.5), ne vom convinge că curbele de nivel sînt în acest cadru niște linii foarte puțin curbate, aproape drepte și paralele. Evident, aceasta este o consecință directă a ipotezei noastre cu privire la regularitatea funcției. Să luăm un punct arbitrar, de exemplu pe linia de nivel 5, și să ducem un cerc cu centrul în acest punct. Diametrul său va fi un segment al curbei de nivel. El împarte cercul în două părți: într-una din ele, notată în figură cu I, toate punctele se află mai sus (nu mai jos) decît cel inițial, în



Fip. 6.5

cealaltă, notată cu II, punctele se află mai jos (nu mai sus!). Să presupunem că punctul marcat este punctul curent de stopare a procesului căutării. În acest caz, firește că înaintarea după toate direcțiile din punctul inițial spre semicercul I conduce la creșterea valorilor funcției-scop, cel puțin în limitele micii porțiuni demarcate. În mod firesc, aceste direcții se numesc *adecvate* sau *convenabile*. E clar de asemenea că între direcțiile convenabile există una cea mai avantajoasă, aceea care pentru o lungime dată a pasului (adică pentru o rază dată) conduce la creșterea cea mai mare a valorii funcției-scop. Această direcție este perpendiculară pe liniile de nivel. Rămîne însă să ne convingem că ea este tocmai direcția gradientului. Pentru comoditate, vom face uz de o nouă figură.

În punctul inițial O vom da coordonatelor x_1 și x_2 asemenea creșteri Δ_1 și Δ_2 încît punctele A și B cu coordonatele $(x_1^{(0)} + \Delta_1,\ x_2^{(0)})$, respectiv $(x_1^{(0)},\ x_2^{(0)} + \Delta_2)$, să se afle pe o aceeași linie de nivel situată cu α mai sus de cea inițială. Înclinarea relativă la x_1 va fi α/Δ_1, iar cea relativă la x_2 va fi α/Δ_2. Acum să luăm pe direcțiile axelor de coordonate segmentele OC și OD proporționale cu înclinările aflate. Atunci segmentul OE va da direcția gradientului. Din asemănarea triunghiurilor OAB și OCE rezultă de îndată că OE este perpendicular pe liniile de nivel.

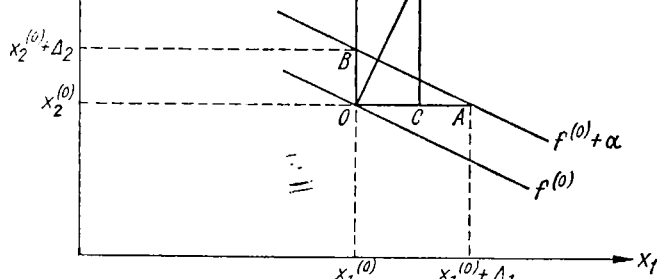


Fig. 6.6

Astfel, pentru cazul intuitiv accesibil care este cazul bidimensional am dovedit că gradientul este perpendicular pe liniile de nivel și deci reprezintă direcția cea mai avantajoasă, cel puțin în limitele unei vecinătăți mici, unde liniile de nivel rămîn paralele.

Firește, cînd pasul este mare, afirmația că direcția gradientului este cea mai avantajoasă devine falsă: în figura 6.7 pasul OB conduce către un punct mai bun decît pasul OC de aceeași lungime, care este efectuat după direcția gradientului OE. Dar studiul proprietăților funcției pentru abaterile mici ale coordonatelor, care rezidă în aprecierea înclinării după fiecare direcție, nu ne putea da, ca atare, alte informații.

Gradientul poate fi prezentat și altfel: dacă vom considera înclinarea după fiecare coordonată ca o forță care tinde să deplaseze punctul respectiv către maximum, atunci gradientul va fi rezultanta tuturor acestor forțe, așa cum se obține aceasta după regulile clasice ale adunării forțelor. Este lesne să ne convingem de aceasta inspectînd figura 6.6, consacrată problemei bidimensionale.

Așadar, înaintînd din punctul inițial în direcția gradientului („și asta-i bine"!), ne-am lovit de valoarea cea mai mare a funcției, corespunzătoare lui $\theta = \theta^{(0)}$, și ne-am

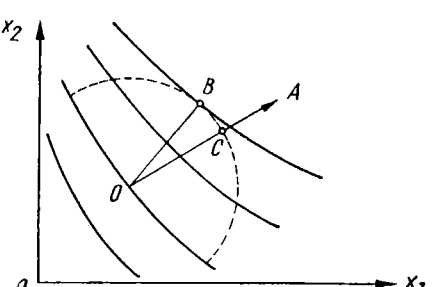


Fig. 6.7

oprit, pentru a chibzui ce mai avem de făcut în punctul cu coordonatele

$$x_1^{(1)} = x_1^{(0)} + \theta^{(0)} \left(\frac{\partial f}{\partial x_1}\right)_0$$

. . . . . .

$$x_n^{(1)} = x_n^{(0)} + \theta^{(0)} \left(\frac{\partial f}{\partial x_n}\right)_0$$

Dar mai departe? Mai departe, ca în basmul cu cocoșul roșu, se repetă totul de la început. Iarăși trebuie să cercetăm o mică vecinătate sau, într-un mod oarecare, să estimăm sau să calculăm derivatele parțiale, să aflăm noua direcție, iar apoi să înaintăm de-a lungul ei pînă la succesul final.

Dar cum vom ști că succesul a fost obținut? Firește că aceasta se va întîmpla tocmai atunci cînd la stoparea curentă valorile derivatelor după toate coordonatele se vor anula, adică va fi imposibil să aflăm în ce direcție să înaintăm în continuare.

Ca și în cazul unidimensional, se poate arăta că punctul în care toate derivatele parțiale se anulează este fie un maximum local, fie un minimum local, fie un punct izolat de inflexiune, fie un punct situat într-un interval în care funcția este constantă. *Fără a încerca demonstrații matematice, să remarcăm că procedura care urmărește să ne conducă spre un minim nu ne poate aduce la un maxim, dacă, firește, cu totul întîmplător, punctul de maxim nu a fost cel inițial. Tot astfel nu vom nimeri într-un punct de minim atunci cînd ne îndreptăm spre un maxim. Pe un interval unde toate punctele sînt puncte de inflexiune, stoparea este posibilă, dar, ca și în cazul unidimensional, ea este practic echivalentă cu pericolul de a ne înfunda într-un extremum local puțin pronunțat.* Deosebirea dintre estimarea derivatelor și valorile lor reale este pe de o parte un necaz, deoarece punctul unde se anulează estimația nu coincide cu extremumul local, iar pe de altă parte utilizarea estimațiilor ne permite să nu observăm micile movile și gropi, așa cum era și în cazul unidimensional.

Cititorul sceptic are, desigur, dreptul să se întrebe **dacă în genere punctul de stopare este accesibil.** Firește că îndoielile sînt justificate, cu atît mai mult cu cît raționamentul n-a fost de fel riguros. Dar, prin analogie cu cazul unidimensional, putem presupune că:

a) dacă parametrul θ este constant, dar nu prea mare, punctele intermediare de oprire se vor apropia de extremum, deși sosirea exactă în punctul respectiv poate să nu se realizeze într-un număr finit de pași;

b) dacă parametrul θ este constant și mare, este posibilă și îndepărtarea de extremum sau oscilarea la mare distanță de el.

A alege între încetineala lui Negrăbilă și temeritatea lui Sarebine este destul de complicat. Din fericire, campionul nostru, Panta-cea-mai-rapidă este în principiu lipsit de asemenea deficiențe. Într-adevăr, pornind dintr-un punct oarecare, înaintarea are loc numai cîtă vreme valorile funcției cresc, și anume pînă la maximum în direcția respectivă, astfel încît valoarea în punctul următor trebuie să fie mai bună decît cea precedentă (procesul conduce la o creștere monotonă, respectiv la o descreștere monotonă atunci cînd se caută minimul). E drept că și aici sînt posibile unele necazuri, dar despre ele vom mai vorbi.

Deocamdată putem constata că am găsit o metodă rațională de a înfrunta multidimensionalitatea. Desigur, metoda are un caracter local; ea constituie, totuși, o mare realizare.

În încheiere, să calculăm un mic exemplu. Funcția-scop fiind dată de formula destul de simplă

$$f(x_1, x_2) = x_1^2 + x_2^2 + 1{,}2x_1x_2,$$

se caută minimul funcției f.

Mai întîi de toate să găsim formulele pentru derivatele parțiale după fiecare coordonată. Amintim că derivata parțială după x_1 se poate calcula ca și derivata obișnuită, dacă se consideră că x_2 este o mărime constantă. Or, în

acest caz f va fi funcție de un singur argument, x_1, și anume un tip bine cunoscut

$$f = x_1^2 + a_1 x_1 + a_0,$$

unde $a_1 = 1,2x_2$, $a_0 = x_2^2$ sînt niște constante.

Amintindu-ne formula din cap. 5, obținem:

$$\frac{\partial f}{\partial x_1} = 2x_1 + a_1 = 2x_1 + 1,2x_2.$$

Printr-un raționament întru totul asemănător, stabilim și formula pentru derivata după x_2:

$$\frac{\partial f}{\partial x_2} = 2x_2 + 1,2x_1;$$

dealtfel, aceasta rezultă și din simetria expresiei $f(x_1, x_2)$*.

De fapt, mai departe aproape că nu mai avem ce calcula. După cum știm, în punctul de minim trebuie să fie satisfăcute condițiile

$$\frac{\partial f}{\partial x_1} = 0, \quad \frac{\partial f}{\partial x_2} = 0.$$

Întrucît în cazul de față dispunem de formule pentru derivate, putem transcrie aceste condiții sub forma unui sistem de ecuații:

$$2x_1 + 1,2x_2 = 0,$$

$$2x_2 + 1,2x_1 = 0.$$

Soluția se obține îndată: $x_1 = 0$; $x_2 = 0$. Acestea sînt coordonatele minimului căutat, în care funcția-scop se anulează.

În principiu, punctele de maxim trebuie și ele să satisfacă aceeași condiție. Dar, evident, luînd x_1, x_2 pozitive și mărindu-le, vom putea

---

* Expresia se numește simetrică atunci cînd valoarea ei nu se schimbă în cazul înlocuirii argumentului x_1 cu argumentul x_2 și reciproc; cu alte cuvinte, dacă $f(x_1, x_2) = f(x_2, x_1)$.

obține pentru $f(x_1, x_2)$ o valoare oricît de mare și, cel puțin, evident mai mare decît zero. Deci valoarea nulă a lui f nu poate fi un maxim. Se poate demonstra, de asemenea, că $x_1 = 0$, $x_2 = 0$ este unicul minim local.

Pentru a convinge pe cititor că această demonstrație, prin urmare și calea analitică directă, nu sînt atît de simple cum s-ar părea, să modificăm întrucîtva funcția-scop. Fie

$$f(x_1, x_2) = x_1^2 + x_2^2 + 3x_1x_2.$$

Nu s-a modificat mai nimic; repetarea cuvînt cu cuvînt a raționamentului de mai sus ne va conduce la concluzia anterioară: punctul de coordonate $x_1 = x_2 = 0$ este un punct de minim. În realitate însă, aceasta nu-i de loc adevărat: în adevăr, funcția-scop considerată se poate transcrie și astfel:

$$f(x_1, x_2) = (x_1 + x_2)^2 + x_1x_2$$

Să considerăm acum punctul unde $x_1 = -x_2 = a$. În acest punct $f(a, -a) = -a^2$ și, prin urmare, pentru orice a diferit de zero valoarea funcției va fi mai mică decît zero. Așadar, $f(0, 0) = 0$ nu este valoare minimă. Mai mult, este clar că minimul funcției-scop poate fi aici atins numai pe frontiera domeniului de căutare. Eroarea rezultatului inițial se explică prin aceea că n-am ținut seama de posibilitatea apariției unor puncte staționare care nu sînt nici puncte de maxim, nici puncte de minim. O examinare mai minuțioasă arată că în fapt punctul $x_1 = x_2 = 0$ este un *punct șa* care este analog cu punctul de inflexiune din problemele unidimensionale (fig. 6.8).

Firește că exemplul simplu considerat este o raritate. În mod obișnuit rezolvarea unui asemenea sistem de ecuații nu este mai lesnicioasă decît aceea a problemei inițiale. Chiar și în cazul unidimensional este adesea nevoie să se facă apel la proceduri de iterare pentru aflarea rădăcinilor; în cazul multidimensional, aceasta este mai curînd regula decît excepția.

De aceea exemplul considerat trebuie privit ca un simplu exercițiu; să încercăm să clarificăm cum îi face față Panta-cea-mai-rapidă.

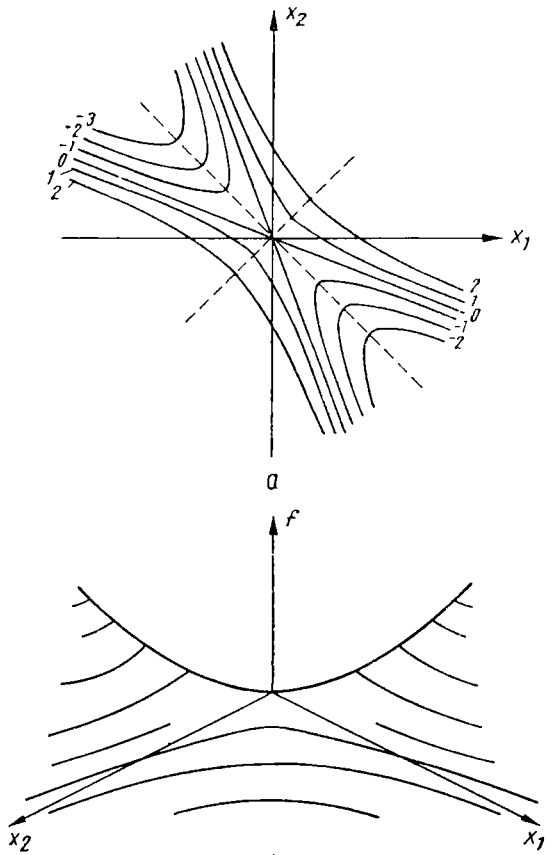


Fig. 6.8

Să pornim de la un punct inițial arbitrar; de exemplu, $x_1^{(0)} = 1$, $x_2^{(0)} = 2$, deci $f(x_1^{(0)}, x_2^{(0)}) = 7{,}4$. Să calculăm* valorile derivatelor parțiale în acest punct; avem:

$$\left(\frac{\partial f}{\partial x_1}\right)_0 = 4{,}4; \quad \left(\frac{\partial f}{\partial x_2}\right)_0 = 5{,}2.$$

Punctul următor trebuie căutat înaintînd pe direcția antigradientului, adică urmărind valorile funcției-scop atunci cînd coordonatele variază după regula

$$x_1(\theta) = x_1^{(0)} - \theta \left(\frac{\partial f}{\partial x_1}\right)_0 = 1 - 4{,}4\theta$$

$$x_2(\theta) = x_2^{(0)} - \theta \left(\frac{\partial f}{\partial x_2}\right)_0 = 2 - 5{,}2\theta$$

unde θ crește pornind de la zero.

Introducînd aceste expresii în formula lui f, obținem

$$f[x_1(\theta), x_2(\theta)] = (1 - 4{,}4\theta)^2 + (2 - 5{,}2\theta)^2 + \\ + 1{,}2(1 - 4{,}4\theta)(2 - 5{,}2\theta),$$

valoare pe care pentru prescurtare o vom nota $F_0(\theta)$. Este clar că această notație este adecvată, întrucît înaintarea are loc după dreapta aleasă; minimul funcției f după această direcție va fi minimul lui F_0 pentru valorile pozitive ale lui θ. După deschiderea parantezelor și reducerea

---

* Dacă nu am ști sau n-am putea calcula direct derivatele parțial, ne-am putea limita la o estimare. Să fixăm mărimea pasului de încercare $\Delta_1 = 0{,}1$ și să calculăm valoarea f; vom obține $f(1{,}1; 2) = 7{,}85$. Atunci estimația derivatei parțiale după x_1, adică înclinarea medie după x_1 se va afla după formula

$$\frac{1}{\Delta_1}[f(x_1^{(0)} + \Delta_1, x_2^{(0)}) - f(x_1^{(0)}, x_2^{(0)})] = \frac{1}{0{,}1}[7{,}85 - 7{,}4] = 4{,}5,$$

ceea ce nu diferă decît puțin de înclinarea adevărată, care este 4,4.

termenilor asemenea, F_0 capătă forma cunoscută a polinomului de gradul doi *

$$F_0(\theta) = 82{,}9\theta^2 - 46{,}4\theta + 7{,}4.$$

Maximul se va afla din condiția

$$F_0'(\theta) = 2 \cdot 82{,}9\theta - 46{,}4 = 0,$$

de unde rezultă că $\theta_{max} = 0{,}280$.

Prin aceasta am determinat cea mai bună valoare a coeficientului θ în pasul inițial. Punînd $\theta^{(0)} = 0{,}280$, putem acum să aflăm și coordonatele punctului de oprire următor:

$$x_1^{(1)} = x_1(\theta^{(0)}) = x_1^{(0)} - \theta^{(0)} \left(\frac{\partial f}{\partial x_1}\right)_0 =$$

$$= 1 - 4{,}4\,\theta^{(0)} = -0{,}230,$$

$$x_2^{(1)} = x_2(\theta^{(0)}) = x_2^{(0)} - \theta^{(0)} \left(\frac{\partial f}{\partial x_2}\right)_0 = 2 - 5{,}2\,\theta^{(0)} = 0{,}543.$$

Dintr-un singur pas ne-am apropiat mult de minim, reducînd valoarea funcției-scop de 15 ori, deoarece $f(x_1^{(1)}, x_2^{(1)}) = 0{,}497$. Al doilea pas îl facem pornind din punctul $x_1^{(1)}$, $x_2^{(1)}$ și urmînd aceeași schemă. Fără a mai da explicații suplimentare, prezentăm rezultatele calculelor succesive:

$$\left(\frac{\partial f}{\partial x_1}\right)_1 = -2 \cdot 0{,}230 + 1{,}2 \cdot 0{,}543 = -0{,}192;$$

$$\left(\frac{\partial f}{\partial x_2}\right)_1 = 2 \cdot 0{,}543 - 1{,}2 \cdot 0{,}230 = 0{,}808;$$

$$x_1(\theta) = -0{,}230 + 0{,}192\,\theta;\quad x_2(\theta) = 0{,}543 - 0{,}808\,\theta;$$

$$F_1(\theta) = (-0{,}230 + 0{,}192\,\theta)^2 + (0{,}543 - 0{,}808\,\theta)^2 +$$

$$+ 1{,}2(-0{,}230 + 0{,}192\,\theta)(0{,}543 - 0{,}808\,\theta) =$$

$$= 0{,}502\,\theta^2 - 0{,}628\,\theta + 0{,}497;$$

* Aici toate calculele se **fac cu trei zecimale exacte.**

$^{(1)} = \theta_{mix} = 0{,}627;$

$\Delta x_1^{(2)} = -0{,}230 + 0{,}192 - 0{,}627 = -0{,}110;$

$x_2^{(2)} = 0{,}038;$ deci $f(x_1^{(2)}, x_2^{(2)}) = 0{,}0085.$

După doi pași, valoarea funcției-scop s-a redus de 870 de ori, iar distanța pînă la minimul adevărat s-a redus aproape de 20 de ori.

Să ne mulțumim cu aceste două iterări și să încercăm să tragem unele concluzii. **În primul rînd,** *în exemplul nostru valorile funcției-scop descresc foarte rapid — vizibil mai rapid decît distanța pînă la extremum. Autorul se grăbește să-l încredințeze pe cititor că acesta este fenomenul obișnuit cînd se utilizează schema pantei celei mai rapide.* **În al doilea rînd,** *lucrul cel mai dificil a fost să calculăm pasul după direcția aleasă.* Aceasta nu mai este o regulă, ci mai curînd e o consecință a faptului că ne-am ocupat cu o problemă bidimensională, în cazul căreia nu e mare lucru să calculezi derivatele. Într-o problemă „cu adevărat multidimensională", aceste calcule se complică; *alegerea parametrului* θ *rămîne totuși, ca și mai înainte, o problemă unidimensională.* În treacăt fie zis, aici apare și „reversul medaliei": *dacă se face uz de căutarea obișnuită după gradient și cu pas constant, dar mic, necesitatea calculării lui* θ *dispare. Numărul iterărilor s-ar putea să crească, dar fiecare iterare în parte va fi mai simplă.* **În sfîrșit,** *cititorul poate obiecta că în genere pentru o problemă atît de simplă toate aceste calcule sînt mult prea greoaie. Dar la aceasta se poate răspunde precum urmează: da, calculele sînt laborioase, dar de unul și același tip și deci comode pentru calculatoare.*

7
Genealogia căutării

(Scheme cu relaxare. Compararea metodelor)

Ce dulce-i murmurul sfioasei fete...
E. Parni

Printre calităţile pe care trebuie să le posede un om de ştiinţă sau un inventator, un rol important îi revine unei atitudini juste faţă de critici. Numai un tînăr naiv îşi închipuie că îndată ce va inventa o nouă maşină, un nou aparat sau chiar o nouă teorie va şi putea începe să se înfrupte din roadele succesului. În afară de faptul că trebuie să ştii să inventezi — şi tocmai aceasta o uită destul de des inventatorii începători — trebuie, de asemenea, să înveţi să aperi dreptul la existenţă a ceea ce ai inventat. Cititorul cult cunoscător în teatru şi literatură îşi va reaminti de o întreagă galerie de tipuri de Inovatori buni şi inteligenţi în luptă împotriva Conservatorilor răi şi proşti, în numele Progresului şi al Demnităţii umane!* Dar noi vorbim despre cu totul altceva. Orice invenţie trebuie supusă unei verificări critice: în primul rînd tre-

* Autorul n-are de gînd să nege că asemenea tipuri există în viaţă. Unul din foştii săi şefi recunoştea: „Ştii cum deosebesc eu ceea ce este realmente nou de ceea ce se preface doar a fi nou? Printr-o metodă pe deplin ştiinţifică: ceea ce este realmente nou va străbate orice barieră. De aceea, încerc să înăbuş, să strivesc tot ce nu este vechi. Dacă ceva reuşeşte să mă învingă, atunci este clar: este ceva realmente nou, nu o falsă inovaţie".

buie lămurit dacă funcționează sau nu: în al doilea rînd trebuie văzut dacă ea diferă cu adevărat de mașinile, aparatele, teoriile cunoscute pînă atunci în domeniul respectiv. Și, în sfîrșit, dacă s-a obținut un răspuns pozitiv la aceste două întrebări, rămîne întrebarea cea mai complicată, principală: oare nu funcționează mai prost decît ceea ce este deja cunoscut? Nici dacă s-a constatat că invenția funcționează mai bine și cu mai mare randament, nici atunci dubiile n-au încetat: căci se poate întîmpla ca ea să fie mai complicată, să fie mult mai scumpă.

Dar cum să învățăm să ne apărăm invenția în condiții atît de grele? Metoda cea mai bună este o metodă simplă: înainte de a aduce prețioasa invenție în fața lumii care o va judeca, trebuie să încercăm singuri să răspundem la întrebările care le-ar putea formula diverși critici. Doar puțini inventatori urmează această regulă utilă, ceea ce, firește, nu-i reduce importanța.

Așadar, să ne închipuim că după îndelungate chinuri de creație am inventat *metoda căutării multidimensionale bazată pe înaintarea după gradient și, în particular, panta cea mai rapidă.* Fiind inventatori raționali, să încercăm să supunem invenția unei critici aspre.

Mai întîi: funcționează? Se pare că da. Am arătat mai înainte considerentele teoretice generale și am procedat la testarea practică a unui mic model (bidimensional). În această privință, totul este în ordine. Problema a doua, a priorității, o vom ocoli — lucru firesc, căci după cum se știe perspectiva istorică nu rămîne neschimbată.

Despre problema a treia însă — cea mai importantă — trebuie să mai vorbim. E drept, pentru a discuta această temă trebuie să știm mai multe despre alte metode, iar noi am făcut cunoștință doar cu un număr foarte mic. *Superioritatea metodei care ne interesează asupra trierii directe (totale)* și asupra înaintării pe o rețea dată în prealabil este clară. Dar nu există oare și alți concurenți?*

* Să amintim că este vorba numai de căutarea unui extremum local! În caz contrar n-avem dreptul să vorbim despre superioritate în raport cu trierea.

Vă rog să vă încredeți în experiența autorului: *există, și încă foarte periculoși.* Cititorul încă n-a făcut cunoștință cu ei, dar, posedînd o imaginație vie, desigur că ghicește cam cum ar fi aceștia.

În primul rînd, oare procedăm rațional în schemele gradientului cînd cheltuim atîta timp pentru încercarea întregii vecinătăți, pentru a măsura înclinarea după fiecare direcție?!

Să analizăm critic această chestiune. De ce am avut nevoie de cercetarea vecinătății? În primul rînd pentru a lămuri dacă nu cumva am ajuns într-un extremum local, iar în al doilea rînd — dacă aceasta nu s-a confirmat — pentru alegerea direcției după care trebuie să înaintăm. Pentru rezolvarea celei de-a doua sarcini, estimăm înclinarea după toate direcțiile alese — axele de coordonate —, iar apoi înaintăm după direcția gradientului (respectiv antigradientului), alegînd coeficientul de proporționalitate θ constant pentru toate coordonatele.

Dacă am alege un coeficient care nu este constant, am înainta nu după gradient (fig. 7.1), ci preferînd una dintre coordonate (fig. 7.1, coordonata x_2). Dacă însă ne gîndim bine, ne dăm seama că „identitatea" coeficientului θ în schema gradientului este o chestiune absolut convențională. Într-adevăr, despre niște θ identici putem vorbi

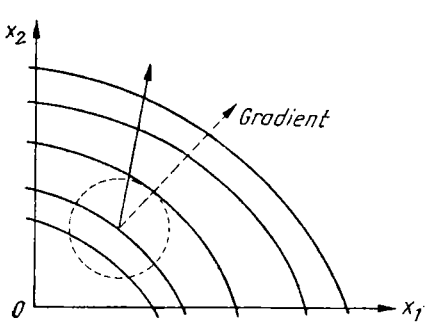


Fig. 7.1

numai dacă toate variabilele se măsoară cu aceeași unitate de măsură, de exemplu în metri, așa cum s-a întîmplat în cadrul „competițiilor" organizate în capitolul precedent, sau dacă toate variabilele se măsoară cu ajutorul numerelor adimensionale, relative, așa cum am convenit în capitolul de față. În caz contrar, formulele pur și simplu nu vor avea sens.

Dar dacă utilizăm mărimi relative (și sîntem nevoiți s-o facem întotdeauna cînd funcția-scop depinde de argumente de natură diferită!)? Atunci direcția în spațiul real al mărimilor inițiale va depinde de scară. Or, după cum se știe, scara este ceva cu totul arbitrar: rezultă că *și o noțiune care la prima vedere pare atît de riguros matematică, noțiunea de gradient, este de fapt destul de arbitrară.*

Să lămurim toate acestea examinînd un caz simplu, bidimensional. Să presupunem, din nou, că se optimizează funcționarea unui cuptor, mai precis se caută o asemenea îmbinare a temperaturii x_1 și a volumului de aer injectat în cuptor x_2 care să asigure randamentul f cel mai bun. Dacă x_1 se măsoară în grade Celsius, iar x_2 în metri cubi pe minut și f în tone pe oră, atunci formulele

$$x_1^{(t+1)} = x_1^{(t)} + \theta \left(\frac{\partial f}{\partial x_1}\right)_t$$

$$x_2^{(t+1)} = x_2^{(t)} + \theta \left(\frac{\partial f}{\partial x_2}\right)_t$$

nu au sens dacă θ este același în ambele formule. Într-adevăr, nu putem aduna decît mărimi de aceeași dimensiune; este un nonsens să adunăm două tone și un grad. Pentru ca să nu se întîmple astfel, să scriem, așa cum se face în fizică, numai dimensiunile și nu mărimile care apar în formule. Obținem:

$$[°C] = [°C] + \theta \left[\frac{\text{t/oră}}{°C}\right],$$

$$\left[\frac{m^3}{min}\right] = \left[\frac{m^3}{min}\right] + \theta \left[\frac{\text{t/oră}}{m^3/min}\right],$$

unde am lăsat θ fără indicarea dimensiunii, dar pentru ca prima formulă să aibă sens trebuie acceptat că θ are dimensiunea $\left[\frac{°C^2 \cdot oră}{t}\right]$, astfel încît, după cum se și cuvine, gradele se vor aduna cu grade. Dar pentru ca a doua formulă să aibă sens, trebuie să presupunem că dimensiunea lui θ este $\left[\frac{m^6 \cdot oră}{t\,(min)^2}\right]$*. Așadar, rezultă că una și aceeași mărime trebuie să aibă dimensiuni diferite, ceea ce este absurd.

Dacă introducem mărimi adimensionale, firește că terminăm cu asemenea neplăceri. Amintim că mărimile relative se introduceau prin formulele

$$\bar{x}_1 = \frac{x_1 - x_{1\,inf}}{x_{1\,sup} - x_{1\,inf}}; \quad \bar{x}_2 = \frac{x_2 - x_{2\,inf}}{x_{2\,sup} - x_{2\,inf}};$$

dar se pot utiliza și alte reguli, de exemplu

$$\bar{x}_1 = \frac{x_1 - x_{1\,inf}}{x_{1\,sup}} \text{ sau } \bar{x}_1 = \frac{x_1}{x_{1\,sup}} \text{ sau } \bar{x}_1 =$$

$$= \frac{x_1}{\frac{1}{2}(x_{1\,sup} + x_{1\,inf})} \text{ etc.}$$

Fiecare din aceste reguli conduce la mărimi adimensionale, toate sînt reguli rezonabile și chiar sînt utilizate în practică, dar, evident, ele conduc la diferite valori ale mărimii relative $\bar{x}_1$, pornind de la una și aceeași mărime inițială x_1. Mai mult, chiar și în prima formulă considerată totul depinde de marginea superioară și de marginea inferioară, care, la rîndul lor, practic sînt determinate mai mult sau mai puțin arbitrar.

Desigur, semnificația rezultatelor definitive nu suferă din cauza acestui arbitrar; odată ce am ales o anumită regulă de transformare a mărimilor absolute în mărimi

* Firește, orele se pot traduce în minute, după cum se întîmplă cu orice măsuri omogene, dar aici aceasta n-ar ajuta la nimic.

relative, putem utiliza oricînd această regulă pentru a efectua transformarea în sens invers.

Formulele pentru determinarea direcției înaintării de la punctul inițial $(x_1^{(0)}, x_2^{(0)})$ către un punct nou sînt cunoscute;

$$\bar{x}_1(\theta) = \bar{x}_1^{(0)} + \theta \left(\frac{\partial \bar{f}}{\partial \bar{x}_1}\right)_0 ; \bar{x}_2(\theta) = \bar{x}_2^{(0)} + \theta \left(\frac{\partial \bar{f}}{\partial \bar{x}_2}\right)_0$$

Ele sînt exprimate, așa cum e și firesc, în coordonate relative. Să arătăm ce vor oferi ele în două cazuri: a) atunci cînd toate mărimile relative sînt obținute din cele absolute după formulele de genul

$$\bar{y} = \frac{y - y_{inf}}{y_{sup} - y_{inf}}$$

și b) atunci cînd se face apel la formule de genul $\bar{y} = \frac{y - y_{inf}}{y_{sup}}$, unde în ambele cazuri locul lui y poate fi ocupat de fiecare dintre literele x_1, x_2 și f. Introducînd aceste definiții în formulele schemei gradientului *, obținem în cazul (a):

$$\frac{x_1(\theta) - x_{1\,inf}}{x_{1\,sup} - x_{1\,inf}} = \frac{x_1^{(0)} - x_{1\,inf}}{x_{1\,sup} - x_{1\,inf}} + \theta \left(\frac{\partial f}{\partial x_1}\right)_0 \frac{x_{1\,sup} - x_{1\,inf}}{f_{sup} - f_{inf}}$$

$$\frac{x_2(\theta) - x_{2\,inf}}{x_{2\,sup} - x_{2\,inf}} = \frac{x_2^{(0)} - x_{2\,inf}}{x_{2\,sup} - x_{2\,inf}} + \theta \left(\frac{\partial f}{\partial x_2}\right)_0 \frac{x_{2\,sup} - x_{2\,inf}}{f_{sup} - f_{inf}}$$

sau:

$$x_1(\theta) = x_1^{(0)} + \theta \frac{(x_{1\,sup} - x_{1\,inf})^2}{f_{sup} - f_{inf}} \left(\frac{\partial f}{\partial x_1}\right)_0$$

$$x_2(\theta) = x_2^{(0)} + \theta \frac{(x_{2\,sup} - x_{2\,inf})^2}{f_{sup} - f_{inf}} \left(\frac{\partial f}{\partial x_2}\right)_0$$

---

* Cînd trecem de la valorile relative ale derivatelor la valorile lor absolute și invers, trebuie să ne amintim că derivatele sînt definite prin intermediul raportului creșterilor.

iar în cazul (b):

$$x_1(\theta) = x_1^{(0)} + \theta \frac{x_{1\,sup}^2}{f_{sup}} \left(\frac{\partial f}{\partial x_1}\right)_0$$

$$x_2(\theta) = x_2^{(0)} + \theta \frac{x_{2\,sup}^2}{f_{sup}} \left(\frac{\partial f}{\partial x_2}\right)_0 .$$

Ne convingem astfel că în formulele pentru mărimile absolute coeficienții derivatelor după prima și a doua coordonată sînt — cum e și firesc — diferiți. După cum am văzut, acest lucru este necesar pentru ca formulele să aibă sens. Dar acum este important altceva: introducerea mărimilor relative prin definițiile (a) respectiv (b) conduce la direcții diferite ale schimbării coordonatelor absolute. Într-adevăr, tangenta unghiului pe care îl determină direcția mișcării cu axa absciselor fiind egală cu raportul creșterilor, în cazul (a) va fi

$$\frac{\left(\frac{\partial f}{\partial x_1}\right)_0 (x_{1\,sup} - x_{1\,inf})^2}{\left(\frac{\partial f}{\partial x_2}\right)_0 (x_{2\,sup} - x_{2\,inf})^2},$$

iar în cazul (b) va fi:

$$\frac{\left(\frac{\partial f}{\partial x_1}\right)_0 x_{1\,sup}^2}{\left(\frac{\partial f}{\partial x_2}\right)_0 x_{2\,sup}^2}.$$

Oricum am alege coeficientul θ, pornind de la un punct dat și înaintînd aparent după unul și același gradient, nu vom putea ajunge într-unul și același loc dacă facem uz (fig. 7.2.) de scări diferite! Așadar, infailibilitatea gradientului — ca a oricărui domnitor absolut — este pusă sub semnul întrebării. *Manipulînd fire ascunse — coeficienții de scară —, putem sili direcția lui să se rotească după voia noastră. Așa încît se poate pune cu bun temei întrebarea: întrucît această direcție este arbitrară,*

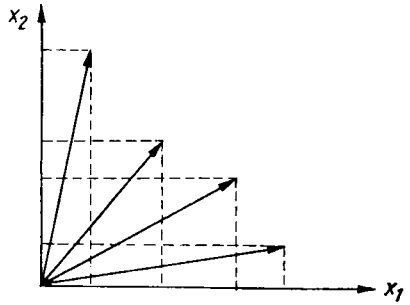


Fig. 7.2

depinzînd de asemenea proceduri care nu decurg direct din probleme cum ar fi metoda alegerii mărimilor relative, mai merită oare să depunem atîtea eforturi pentru determinarea ei?

Şi totuşi, nu putem răspunde fără drept de apel: „nu merită!" Mai întîi de toate, observăm că, deşi alegerea scării influenţează mărimea creşterii, totuşi semnul creşterii rămîne neschimbat şi este pe deplin determinat de către semnul derivatei respective. De aceea, „manipularea gradientului" pe baza schimbării scării poate să aibă loc numai în anumite limite: în nici un caz nu-l putem roti în direcţie opusă. Mai mult, toate direcţiile posibile se încadrează în limitele unui unghi drept (în cazul bidimensional pe care îl reprezintă fig. 7.2), iar în cazul general, n-dimensional, în limitele unui ortant. Aşadar, arbitrarul este limitat.

Şi totuşi din însăşi interpretarea acestui arbitrar se pot trage o mulţime de concluzii utile.

În primul rînd, întrucît arbitrarul este limitat de caracterul determinat al semnului derivatelor, rezultă că şi în cazul „sondării" vecinătăţii putem să ne mărginim numai la estimarea semnului înclinării, adică să nu calculăm creşterile, ci să ne mărginim la comparaţii (mai mare, mai mică). Să amintim că aceasta este suficient şi pentru verificarea caracterului de extremum. Or, este mai simplu să facem comparaţii decît scăderi şi împărţiri, cum era necesar în cazul estimării derivatelor.

De aici decurge caracterul rațional al unui nou algoritm, ba chiar al unei familii întregi de algoritmi de iterare. În cazul căutării maximului, acțiunea lor este următoarea: pornind dintr-un punct arbitrar și nimerind după cel de-al t-lea pas „mare" în punctul de coordonate $(x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)})$, *recurgem la „sondare" comparînd valorile în punctele vecine, iar apoi facem al* $(\mathbf{t}+\mathbf{1})$*-lea pas „mare" către un nou punct, în funcție de rezultatele comparației.*

Întrucît după toate coordonatele se efectuează operații de același tip, să încercăm a exprima formula pentru prima coordonată

$$x_1^{(t+1)} = x_1^{(t)} + \begin{cases} \theta, \text{ dacă } f(x_1^{(t)} + \Delta_1, x_2^{(t)}, \ldots, x_n^{(t)}) > \\ \qquad > f(x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)}), \\ -\theta, \text{ dacă } f(x_1^{(t)} + \Delta_1, x_2^{(t)}, \ldots, x_n^{(t)}) < \\ \qquad < f(x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)}), \\ 0, \text{ dacă } f(x_1^{(t)} + \Delta_1, x_2^{(t)}, \ldots, x_n^{(t)}) = \\ \qquad = f(x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)}). \end{cases}$$

Nici nu e nevoie neapărat să analizăm formula, ajunge doar să privim figura care arată în ce parte trebuie să facă „omul care caută" pasul de mărime θ. Această schemă se numește *schema pasului*, întrucît aici fixînd pe θ determinăm nu un coeficient de proporționalitate, ci chiar mărimea pasului „mare". Pașii în direcția diverselor coordonate firește că pot fi diferiți, dar ei se execută simultan,

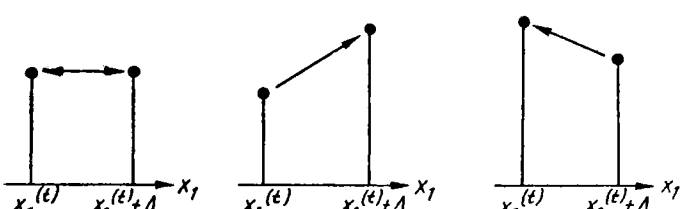


Fig. 7.3

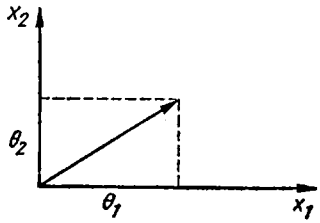


Fig. 7.4

astfel încît, mărimile fiind date, direcția este aleasă (fig. 7.4). Dacă toate mărimile sînt date (fiind egale pentru toți pașii), metoda se numește *metoda pasului constant.* Dacă o examinați cu atenție, vă veți aminti de o veche cunoștință. Într-adevăr, dacă pasul este dat, fiecare axă de coordonate va fi divizată în segmente egale între ele; dacă înaintăm în conformitate cu această metodă, ne vom putea opri numai la extremitățile acestor segmente. Or, dacă considerăm toate pozițiile posibile în spațiu sau cel puțin în plan, vom descoperi vechea noastră cunoștință, rețeaua! „Omul care caută" poate păși numai pe nodurile acestei rețele (fig. 7.5).

Dar firește, că schema pasului constant nu coincide cu tactica lui Rețeanu, așa cum am descris-o în cap. 4: într-a-

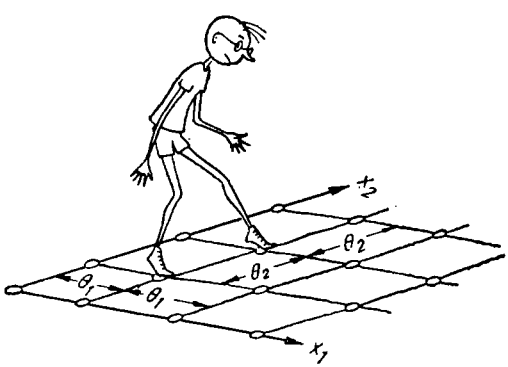


Fig. 7.5

devăr, sondarea se face nu după nodurile rețelei, ea determină situația numai după efectuarea unor pași „mari". Ca și în alte scheme care utilizează ideea de extrapolare, noua schemă nu stabilește o legătură rigidă între „gîndire" și „acțiune". Pe de altă parte, amintirile ne vor ajuta totuși să înțelegem particularitățile schemei noi.

În primul rînd, întrucît numărul nodurilor într-un domeniu mărginit este finit, procesul de iterare după un număr finit de pași ne va aduce în vecinătatea unui extremum. S-ar putea spune că el va stopa, va converge într-un număr finit de pași, dar aceasta ar fi oarecum o imprudență. Există anumite situații neplăcute, cu care cititorul poate face cunoștință chiar pe modelul unidimensional (fig. 7.6). Este evident că ajungînd, după pasul al t-lea, în punctul x și sondînd vecinătatea, „omul care caută" va fi nevoit să facă un pas la dreapta, de mărimea θ, ajungînd la punctul $x + \theta = y$; sondînd aici, va ajunge însă la concluzia că trebuie să facă un pas la stînga și va nimeri iarăși în punctul x!

După cum se spune, aici avem un „ciclu"; „omul care caută", în loc de a se opri locului în vecinătatea extremumului, va fi nevoit să dănțuiască la infinit, sărind mereu de la x la y și înapoi.

Iată deci că am constatat o deficiență importantă a schemei: poate să apară un ciclu. *În plus, dacă fixăm oprirea într-unul din punctele ciclului, de exemplu în x, precizia nimeririi în punctul de extremum va fi de ordinul θ și nu de ordinul Δ, cum se întîmpla în schema gradientului. Evident, mărimea pasului poate fi schimbată după împre-*

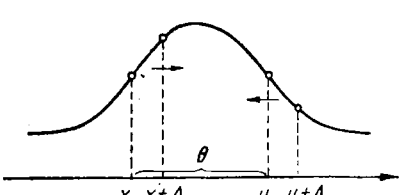


Fig. 7.6

jurări; de exemplu, în etapele inițiale, îndepărtate de extremum, putem adopta un pas mai mare, micșorîndu-l apoi treptat sau micșorîndu-l atunci cînd se constată un ciclu.

Dar, la drept vorbind, **acestea vor fi acum scheme noi, scheme cu pas variabil.** Scheme noi, deci și merite noi și deficiențe noi.

Meritele sînt clare, dar lipsurile? Ele provin din însăși ideea de pas variabil; dacă variezi pasul, trebuie să știi cum să-l variezi. Lucrul cel mai firesc ar fi să-l variezi potrivit cu înclinarea: dacă înclinarea este mai mare, atunci și pasul este mai mare. Dați-mi voie — va obiecta cititorul indignat —, în felul acesta reveniți la schema gradientului, pe care ați criticat-o; din nou va trebui să estimați mărimile derivatelor!

Și, incontestabil, cititorul va avea dreptate: critica noastră a efectuat un cerc închis, am ajuns de unde am pornit. *Pentru orice metodă care se află pe acest drum există cîte un pro și cîte un contra corespunzător și nu putem acorda o preferință decisivă vreuneia dintre ele.* Dar în acest caz, mai merita oare să ne ocupăm cu critica? — va spune cititorul sceptic. Desigur că da, merita — îi vom răspunde noi; critica a fost cu adevărat rodnică; ea a generat o întreagă familie de noi proceduri de căutare.

Dealtfel, nici nu am făcut încă cunoștință cu toți membrii acestei familii. Există și o „subfamilie" rațională: ca și în cazul pantei celei mai rapide, alegînd direcția putem să adaptăm lungimea pasului, parametrul θ, înaintînd pînă la extremum după direcția aleasă, și să efectuăm căutarea unidimensională a extremumului prin diferite proceduri. În practică, toți membrii acestei „subfamilii" a schemelor cu iterare funcționează destul de bine și, firește, sînt mai efectivi decît strămoșul lor, schema simplă cu pas constant.

Dar autorul este grăbit să vă facă cunoștință cu o altă familie, mai veche și mai renumită, care de mult concurează cu familia gradientului pentru supremația în sînul marelui trib al procedurilor de căutare iterată — și, după cum ne vom convinge, are tot dreptul s-o concureze. Numele acestei familii sună neobișnuit și e

destul de greu de înțeles: *metoda relaxării*. Originea numelui este realmente veche și încercarea de a-l explica ne-ar duce departe de cadrul narațiunii noastre, introducîndu-ne în labirintul teoriei lucrărilor de construcții. Dealtfel, *în traducere cuvîntul „relaxare" înseamnă pur și simplu slăbire a unei legături, a unei tensiuni. După cum vom vedea mai departe, denumirea este justificată.*

Să lăsăm însă la o parte toate aceste considerente și să trecem la fondul chestiunii, ca să aflăm ce se ascunde în spatele acestui cuvînt.

Este un lucru mai mult decît simplu. Într-adevăr, să luăm, ca întotdeauna, un punct inițial arbitrar, cu coordonatele $x_1^{(0)}$, $x_2^{(0)}$, ..., $x_n^{(0)}$, valoarea funcției în acest punct fiind $f^{(0)} = f(x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)})$. Am fixat astfel poziția punctului în spațiu. *Acum să începem a ne elibera de conexiuni, îngăduind coordonatelor să varieze una cîte una. Mai întîi, vom modifica prima dintre ele, x_1, lăsînd neschimbate valorile celorlalte, și vom efectua căutarea unidimensională după x_1, pînă ce vom da de extremum după direcția acestei axe. Acolo, adică în punctul de coordonate $x_1^{(1)}$, $x_2^{(0)}$, $x_3^{(0)}$, ..., $x_n^{(0)}$, ne vom opri și vom relaxa a doua coordonată, permițînd punctului să se miște pe axa respectivă pînă la extremum, celelalte coordonate rămînînd rigide (coordonata x_1 are deja noua valoare $x_1^{(1)}$, obținută în urma pasului precedent!). Această procedură se repetă de n ori, în care timp „relaxăm" pe rînd toate cele n coordonate. Apoi trecem iar la coordonata x_1 și organizăm un nou rînd de n pași și așa mai departe, pînă la stopare. Aceasta va avea loc atunci cînd punctul respectiv va fi punct de extremum după direcțiile tuturor axelor de coordonate**. **Nu este așa că metoda e uimitor de simplă și elegantă? Nu degeaba născocirea ei este legată de numele genialului matematician Gauss.** *Ca și în cazul pantei celei mai rapide,* **prin această metodă căutarea multidimensională se trans-**

* Este evident că dacă funcția-scop reprezintă de fapt o sumă de funcții, astfel încît fiecare din ele este funcție de un singur argument, adică dacă $f(x_1, x_2, ..., x_n) = f_1(x_1) + f_2(x_2) + ... + f_n(x_n)$, extremumul va fi atins într-un singur rînd de n pași.

formă într-o succesiune de căutări unidimensionale, dar fără nici un fel de prospectări în scopul alegerii direcției acestor înaintări unidimensionale, ci doar prin trierea rînd pe rînd a tuturor direcțiilor axelor de coordonate. *Și totuși succesul vine aproape întotdeauna. Simplitatea schemelor cu relaxare (cititorul probabil că-și imaginează variantele posibile pentru o atare schemă) a făcut din ele de multă vreme instrumentul preferat al cercetătorilor. Mai mult, adesea apar noi lucrări științifice ce prezintă noi scheme de căutare, dar, dacă le analizăm cu atenție, iese la iveală aceeași idee a relaxării.*

Avem, așadar, o nouă pasiune. Să renunțăm la obositoarea procedură a aflării gradientului și... să zicem într-un glas: „Nu există alt Dumnezeu în afară de RELAXARE, iar Gauss este profetul ei". Dar... după cum a spus Goethe, marele contemporan al lui Gauss, nu există faptă pentru care să nu trebuiască să plătești. Renunțînd la necesitatea alegerii direcției, dăm de o serie de neplăceri. Mai întîi, o schemă de relaxare poate conține mult mai mulți pași de căutare unidimensională decît schema pantei celei mai rapide. În schemele de relaxare numărul pașilor este foarte sensibil la orientarea liniilor de nivel ale funcției-scop față de axele de coordonate. Pentru a ne convinge de acest lucru, iată un exemplu clasic. Să presupunem că se caută minimul funcției f determinată prin formula $f(x_1, x_2) = a^2x_1^2 + x_2^2$. Se vede ușor că liniile de nivel constant determinate de condițiile $a^2x_1^2 + x_2^2 = c$, unde a și c sînt constante, sînt niște elipse ale căror semiaxe sînt orientate după axele de coordonate (în fig. 7.7 sînt indicate curbele de nivel pentru $a = 1/2$). Este clar că pornind din punctul inițial A, schema cu relaxare ne va conduce în doi pași la minimul $x_1 = 0$, $x_2 = 0$. Să rotim acum toate elipsele cu 45° în raport cu axele sau, ceea ce revine la același lucru, să facem o rotație de 45° a axelor de coordonate (fig. 7.7). După rotire valoarea coordonatelor se va schimba; și amintindu-ne de cele știute din trigonometrie

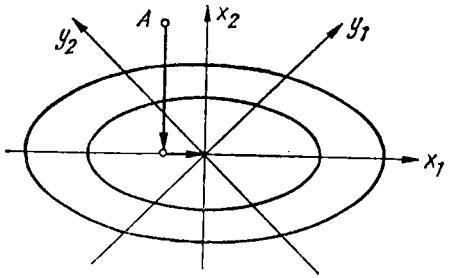


Fig. 7.7

sau, mai simplu, de teorema lui Pitagora (a se vedea fig. 7.8), va fi ușor să ne convingem că punctul cu noile coordonate y_1, y_2 avea anterior coordonatele $x_1 = \frac{y_1 - y_2}{\sqrt{2}}$, $x_2 = \frac{y_1 + y_2}{\sqrt{2}}$.

Funcția-scop rămîne aceeași, dar formula în raport cu noile axe de coordonate va fi diferită:

$$f(y_1, y_2) = \frac{a^2}{2}(y_1 - y_2)^2 + \frac{1}{2}(y_1 + y_2)^2 =$$

$$= \frac{1}{2}(a^2 + 1)\, y_1^2 + \frac{1}{2}(a^2 + 1) y_2^2 - (a^2 - 1) y_1 y_2.$$

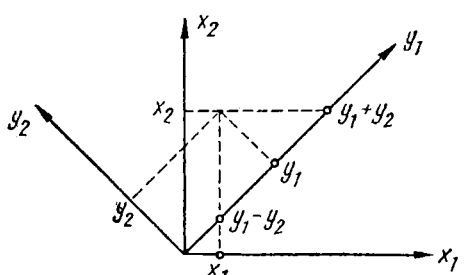


Fig. 7.8

Acum funcția-scop nu mai este o sumă de funcții de un argument și, ca atare, minimul ei nu mai este accesibil atît de ușor.

Într-adevăr, drumul în zigzag pe care îl impune relaxarea este indicat în figura 7.9. Teoretic vorbind, pentru a ajunge exact în punctul de extremum este necesar un număr infinit de pași. Și totuși funcția nu a devenit mai complicată!

În schimb, schema gradientului alege parcă singură, la fiecare pas, direcția axei după care trebuie efectuată înaintarea, iar această axă este de obicei mai apropiată de direcția rezonabilă care conduce de-a dreptul la țel. În finalul capitolului 6 am efectuat calculul pentru o asemenea funcție după metoda pantei celei mai rapide. În numai doi pași nimeriserăm într-o mică vecinătate a extremumului, apropiindu-ne de el cu 0,05 din distanța inițială.

Pe de altă parte, deși în cazul acestor axe de coordonate incomode pentru metoda relaxării a fost necesar un drum complicat, în zigzag și cu un mare număr de iterări pentru a ajunge la minim, merită totuși să analizăm cu atenție cum s-a efectuat calculul.

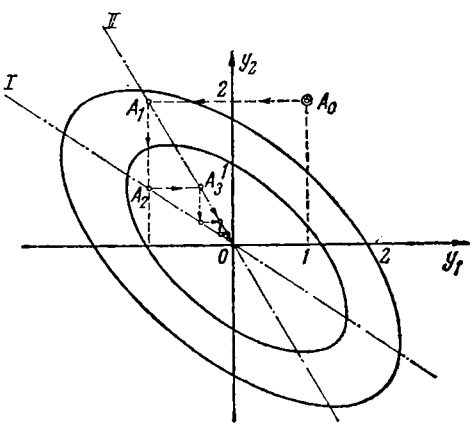


Fig. 7.9

Pentru $a = 1/2$, funcția-scop se scrie în felul următor*:

$$f(y_1, y_2) = y_1^2 + y_2^2 + 1{,}2y_1y_2$$

Căutarea începe din punctul de coordonate $y_1^{(0)} = 1$; $y_2^{(0)} = 2$ (fig. 7.9). Să eliberăm coordonata y_1, y_2 rămînînd rigidă și egală cu $y_2^{(0)} = 2$. Atunci funcția-scop va avea forma

$$f(y_1, y_2^{(0)}) = y_1^2 + (y_2^{(0)})^2 + 1{,}2y_1y_2^{(0)} = y_1^2 + 2{,}4y_1 + 4$$

devenind astfel funcție de un singur argument, y_1. Minimul se află derivînd, adică este dat de condiția

$$\frac{\mathrm{d}}{\mathrm{d}y_1} f(y_1, y_2^{(0)}) = 2y_1 + 1{,}2y_2^{(0)} = 2y_1 + 2{,}4 = 0,$$

de unde rezultă că coordonatele noului punct sînt următoarele:

$$y_1 = y_1^{(1)} = -1{,}2; \quad y_2 = y_2^{(0)} = 2.$$

Acum fixăm coordonata y_1 la acest nou nivel și permitem variația lui y_2. Funcția-scop devine iarăși funcție de un singur argument:

$$f(y_1^{(1)}, y_2) = (y_1^{(1)})^2 + y_2^2 + 1{,}2y_1^{(1)}y_2 = 1{,}44 + y_2^2 - 1{,}44\, y_2.$$

Minimul îl vom afla acum din condiția

$$\frac{\mathrm{d}}{\mathrm{d}y_2} f(y_1^{(1)}, y_2) = 2y_2 + 1{,}2y_1^{(1)} = 2y_2 - 1{,}44 = 0$$

și ajungem astfel la punctul $y_1 = y_1^{(1)} = -1{,}2$; $y_2 = y_2^{(1)} = 0{,}72$.

În continuare, începe ciclul următor, adică revenim la variabila y_1, apoi la y_2 etc. Rezultatele calculelor sînt prezentate în tabel, precum și, după cum am mai spus, în figura 7.9.

---

* De fapt, după substituția $a = 1/2$ în formula precedentă, pentru f se obține o funcție care diferă de cea scrisă cu factorul 5/8. Este însă evident că simpla înmulțire cu un factor constant a formulei care determină funcția-scop nu poate modifica amplasarea extremumului.

După cum se vede, deși numărul iterărilor este mare, fiecare în parte este cu mult mai simplă decît în metoda pantei celei mai rapide; ajungem astfel la binecunoscutul „da, dar", adică există argumente și pro și contra.

Nr.	0	1		2		3		4	
y_1	1,0	−1,20	−1,20	−0,432	−0,432	−0,155	−0,155	−0,056	−0,056
y_2	2,0	2,0	0,72	0,72	0,259	0,259	0,093	0,093	0,033
f	7,4	2,56	0,92	0,09	0,030	0,011	0,0037	0,0013	0,0005

Acum să luăm seama (fig. 7.9) la o proprietate interesantă a traiectoriei căutării în exemplul considerat: punctele iterărilor succesive se află pe razele I și II determinate de ecuațiile

$$y_2 = -\frac{2}{1,2}\, y_1, \text{ respectiv: } y_2 = -\frac{1,2}{2}\, y_1,$$

iar traiectoria coboară treptat, în zigzag, către originea axelor de coordonate între aceste două raze. Dar ce-ar fi dacă, organizînd căutarea, am utiliza de la început această proprietate? Efectuînd primii trei pași, obținem două puncte A_1 și A_3, care se află pe o dreaptă ce merge într-adevăr spre minim. Dacă lucrurile stau așa, rezultă că, efectuînd un singur pas pe această dreaptă, ajungem dintr-o dată la extremum (fig. 7.9). Acum în locul unei scărișoare infinite avem nevoie doar de patru pași. O asemenea variantă a coborîrii se numește adesea *metoda tangentelor paralele (TP)*. Apariția acestei denumiri se explică prin aceea că metoda *TP* este direct sau indirect legată de următoarea proprietate a familiei de elipse concentrice care reprezintă în exemplul nostru liniile de nivel: dacă în punctul inițial A_0 ducem o tangentă la linia de nivel (fig. 7.10), iar apoi ducem o dreaptă oarecare paralelă cu ea, această nouă dreaptă va fi de asemenea tangentă la o curbă de nivel, iar minimul, centrul comun al elipselor se va afla obligatoriu pe dreapta determinată de cele două puncte de tangență.

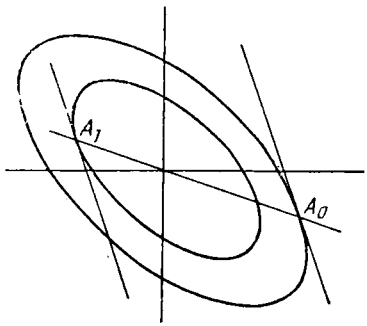


Fig. 7.10

Este evident că dreptele care în figura 7.9 trec prin A_0A_1 și A_2A_3 sînt paralele și tangente în punctele A_1 și A_3 ale curbelor de nivel corespunzătoare. Prin urmare am utilizat tocmai această proprietate a tangentelor paralele. Demonstrația riguroasă a acestei proprietăți și cu atît mai mult extinderea ei la cazul n-dimensional (cînd $n > 2$) este suficient de complicată și o vom omite. Cu metoda TP puteți face cunoștință mai amănunțit după cartea [28]. Ne vom limita la o singură observație. Aici posibilitatea accelerării este legată de caracterul eliptic al curbelor de nivel. Dacă însă în realitate curbele de nivel nu sînt elipse, atunci, în cazul general, metoda poate să-și piardă eficacitatea. Totuși, în practică se constată că, dacă funcția-scop este suficient de regulată în vecinătatea extremumului, curbele de nivel sînt de obicei apropiate de elipse și metoda, deși nu ar conduce către un extremum exact, dă o aproximare destul de bună. Avem aici o analogie deplină, atît ca formă cît și ca fond, cu metoda lui Newton. Caracterul relativ restrîns al metodei TP, „predilecția" sa pentru o anumită clasă de suprafețe era evidentă de la bun început. De aceea să încercăm încă o dată să căutăm concurenți mai norocoși ai pantei celei mai rapide în sînul unor metode mai universale.

8
Spre succes prin legea numerelor mari

(Căutarea locală la întîmplare)

Sub norul cel negru
Pierit-a jumătate de cer.
Gemînd printre pini
Vîntul trage alarma.
Norul cel negru a şi acoperit pămîntul.
Găsi-va oare drumul Peer Gynt?
Gabriela Mistral, *Cîntecele lui Solveig*

Schemele de relaxare constituie fără îndoială o familie însemnată, de mare vază şi cu merite deosebite, care pe drept cuvînt se mîndreşte atît cu vechimea ei cît şi cu succesele ei actuale. Dar oricît de serioasă ar fi în aparenţă comportarea lor, în spatele seriozităţii se ascunde totuşi o anumită superficialitate. Într-adevăr, cum realizează ele alegerea direcţiei? S-ar părea că într-un mod foarte serios: după numărul de ordine al coordonatei. Dar de unde mai apare şi această ordine? De ce temperaturii să-i zicem x_1, iar volumului de aer x_2 şi nu dimpotrivă? De fapt, lucrul acesta se face la întîmplare, fără nici un fel de temei sau regulă de preferinţă. Mai mult, chiar direcţiile axelor sînt de fapt întîmplătoare, întrucît întotdeauna înainte de începerea căutării ar putea avea loc o transformare a axelor de coordonate, care, după cum am şi văzut, poate să îmbunătăţească, dar să şi înrăutăţească starea de lucruri.

Aşadar, **în impunătoarele scheme de relaxare există şi un element ascuns de superficialitate.** E limpede însă că oarecare superficialitate prezintă şi schemele gradientului: în primul rînd, direcţia gradientului depinde de scara stabilită la întîmplare pe fiecare axă, iar în al doilea

rînd prospectarea („sondarea") vecinătății fiecărui punct are loc numai după direcția axelor de coordonate, iar direcțiile sînt alese arbitrar! Firește, toate acestea sînt adevărate și pentru schemele pasului constant.

Ne-am convins astfel, că **deși, potrivit aparențelor, procesele de căutare pe care le-am studiat sînt determinate sau, ca să folosim un limbaj mai savant, sînt caracterizate printr-un strict determinism, este totuși prezent și un element de intuiție, de ghicire, de aleatoriu. Ce-i drept, acest element apare o singură dată, înainte de începutul căutării, dar el influențează întregul mers al căutării.**

Într-o anumită măsură, situația amintește de o partidă de cărți între doi jucători excepționali, care joacă fără cea mai mică greșeală, cu cărțile pe masă. De fapt, jocul lor este predeterminat de cărțile care le-au venit, dar cărțile acestea au fost extrase din pachet la întîmplare și tocmai ele determină rezultatul partidei. Or, de fapt, căutarea este și ea un joc, un joc împotriva Naturii, care face prima mișcare, propunîndu-ne o enigmă dintr-o infinitate: o problemă de căutare *.

Fiecare funcție-scop poate fi considerată ca fiind extrasă prin hazard dintr-un pachet de cărți pe care-l amestecă Natura. Succesul nostru depinde de felul în care ne vom adapta acestei funcții, de priceperea de a alege strategia căutării.

Așadar, în schemele obișnuite, totul este predeterminat de prima alegere întîmplătoare: Natura alege una dintre funcții, noi alegem una dintre metodele căutării. Or, se poate întîmpla ca alegerea noastră să fie nereușită, metoda să fie nepotrivită pentru funcția dată; schemele obișnuite pe care le cunoaștem nu mai sînt în stare să repare greșeala.

De aceea, *este firesc să încercăm a utiliza o strategie mai suplă, introducînd un element de alegere la întîmplare*

---

* În această carte nu dăm nici un fel de informații riguroase cu privire la teoria jocurilor. Cititorul care cunoaște această teorie va putea să înțeleagă mai bine ideea jocului împotriva naturii citind cap. 12.

la fiecare pas. Poate că în acest caz vom reuși să ne adaptăm mai bine la funcția aleasă de Natură.

Să privim cu mai multă atenție schemele studiate. Cu toate diferențele existente, apare totuși o trăsătură comună: în fiecare punct se alege una dintre direcțiile convenabile care formează un unghi ascuțit cu gradientul, apoi are loc căutarea unidimensională după direcția aleasă. Pentru alegerea direcției convenabile se depun eforturi considerabile. Într-o problemă n-dimensională, pentru acest lucru sînt necesare: în schema gradientului $n + 1$ măsurători, precum și prelucrarea lor care comportă n scăderi *; în schemele pasului constant, tot atîtea măsurători și n comparații; în schemele de relaxare trei (sau două) măsurători și o scădere (sau comparare) care determină alegerea direcției (la dreapta sau la stînga) pe axa determinată în prealabil. Schemele gradientului se bazau pe ideea alegerii mai binelui, cel puțin în ce privește pașii mici, celelalte însă se mulțumeau cu alegerea unei direcții convenabile arbitrare, ceea ce limita eforturile în mod corespunzător.

În aceste scheme, fiecare pas poate fi mai puțin reușit, adică poate să corespundă unei creșteri mai mici a valorilor funcției-scop, astfel încît unei înaintări în sus obținută printr-un singur pas al schemei gradientului să-i corespundă mai mulți pași în schemele mai simple; în schimb, eforturile totale depuse pentru înaintarea respectivă pot să fie mai mici.

Așadar, și aceste considerente ne conduc la concluzia că s-ar putea ca alegerea direcției în fiecare punct să nu merite acea atenție deosebită pe care o pretind schemele gradientului.

Poate că am putea s-o alegem la întîmplare, pe ghicite, și anume la fiecare pas independent de ceea ce a avut loc la pașii precedenți?

Tocmai această idee se află la baza unei alte familii din glorioasa seminție a schemelor cu iterare; familia

* Împărțirea cu Δ poate fi omisă, considerînd că această operație este compensată de alegerea coeficientului θ.

metodelor căutării la întîmplare. Este familia cea mai tînără, dar foarte numeroasă și în rapidă dezvoltare.

Vom face cunoștință numai cu unii reprezentanți ai ei, mai cunoscuți.

Să-l prezentăm pe primul. Acesta are o tactică foarte simplă: ca și rudele sale din familia relaxării, el se bazează pe principiul alegerii direcției numai după axele de coordonate, renunțînd însă la condiția: „trecerea la o direcție nouă are loc după o regulă stabilită o dată pentru totdeauna". De fiecare dată această metodă alege direcția în mod întîmplător, trăgînd-o la întîmplare din „pachetul direcțiilor".

Aici fiecare înaintare unidimensională poate avea loc după orice regulă. Dacă se utilizează ideea pantei celei mai rapide, adică a înaintării pînă la extremum după direcția aleasă, atunci firește că posibilitățile unei direcții date se epuizează astfel încît la pasul următor, înainte de a alege, această direcție trebuie eliminată din „pachet".

Dar iată și al doilea reprezentant al familiei. Acesta refuză cu totul moștenirea oferită de metoda relaxării și alege direcția mișcării fără a mai pune restricția ca ea să coincidă cu una din axele de coordonate. Pasul de încercare de lungime Δ se face într-o direcție arbitrară.

Dacă în noul punct A, situat pe sfera de rază Δ și cu centrul în punctul inițial O (fig. 8.1), valoarea funcției-scop va fi mai mare, atunci (dacă se caută maximul) direcția

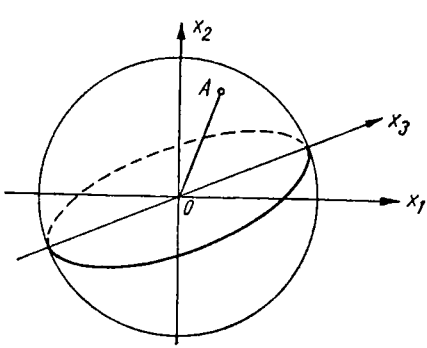


Fig. 8.1

de la *A* la *O* este admisă drept convenabilă pentru căutarea unidimensională. Dacă însă valoarea în *A* este mai mică, atunci se adoptă direcția opusă.

Al treilea membru al familiei are însă opinii de-a dreptul anarhiste. Principiul alegerii direcției cu ajutorul studierii unei mici vecinătăți, alegere urmată de căutarea unidimensională, nu-i convine deloc. Deviza lui e simplă: „Procedează cum îți vine la îndemînă". Dacă iese, e în regulă; dacă nu, nu-i nimic, nu-i mare pagubă, mai repeți: cine știe, poate data viitoare vei avea noroc.

În conformitate cu această deviză, el pornește de la bun început cu un pas mare într-o direcție aleasă la întîmplare. Lungimea pasului de obicei o alege constantă: alteori modifică lungimea de la un pas la altul după o lege dată, dar și pe ea o poate lua la întîmplare. Această metodă este cea mai originală și se deosebește cel mai pregnant de cele pe care le cunoaștem.* Vom încerca s-o studiem, comparînd acțiunea ei fie și cu schema pasului constant în direcția gradientului.

Numai că studiul nu va fi ușor. Într-adevăr, noua metodă nu este pur și simplu una din numeroasele metode care fac parte din tribul proceselor cu iterare; de fapt, ea este venită dintr-o altă lume, tainică și deocamdată întru totul necunoscută; lumea întîmplării, lumea aleatoriului. Această lume își are propriul său cod de legi, denumit *teoria probabilităților*. Este un cod destul de dificil și de complicat și pentru a-l cunoaște cu adevărat, în esența lui, ar fi necesare mult mai mari eforturi decît pentru asimilarea ideilor de bază ale căutării.

De aceea, nici nu ne vom apuca de o treabă atît de grea, ci vom organiza doar o mică excursie în lumea aleatoriului, căci, din moment ce am deschis prima portiță spre această lume, nu vom putea evita nici ulterior întîlnirile cu ea. În timpul excursiei va trebui să facem cunoștință cel

* În acest capitol ne vom interesa numai de asemenea proprietăți ale căutării la întîmplare care se manifestă în cazul unor pași nu prea mari; chiar și alegerea la întîmplare a lungimii va fi supusă unor restricții. Astfel vom avea posibilitatea să ne menținem în cadrul schemelor locale.

puțin cu principalele reguli de comportare în lumea aleatoriului *.

Pentru un străin, această lume este foarte ciudată: „o lume nebună, nebună, nebună". Conduita „locuitorilor" săi pare să nu se supună nici unei norme rezonabile. Ei sînt într-o permanentă agitație, trecînd mereu de la o treabă la alta; la fiece schimbare ei dau fuga mai întîi într-o mică dugheană în care își are reședința o bătrînă ghicitoare, înconjurată de atributele obligatorii ale meseriei sale: zarul, cărțile de joc, roata norocului (ruletă) și celelalte. Pentru a putea înțelege ceva din toate acestea, să încercăm a sta de vorbă cu unul dintre locuitori. Iată-l pe unul care tocmai a ieșit din dugheana ghicitoarei.

Noi: *Scuzați-ne. Am venit într-o excursie și am vrea foarte mult să cunoaștem minunata dv. lume. N-ați putea să ne ajutați în această privință?*

El: *Cu plăcere. Mă numesc Epsilon și sînt la ordinele dv.*

Noi: *Poate că nu prea este cuviincios din partea noastră, dar am dori să vă întrebăm de ce dv. și compatrioții dv., care la prima vedere păreți a fi oameni foarte serioși, alergați mereu la ghicitoare?*

Epsilon *(mirat): Dar cum s-ar putea altfel? Eu, de exemplu, trebuie să decid astăzi ce am de făcut: la prînz, pe masă vor fi și prăjitură, și cremă de ciocolată cu frișcă. Îmi plac tare mult și*

---

* Expunerea noțiunilor de bază ale teoriei probabilităților pornind de la matematica elementară, ba încă pe spațiul limitat a zece pagini, este, firește, o treabă dificilă, dacă nu chiar disperată; dar conceptele probabiliste sînt mult prea importante pentru teoria căutării pentru a nu spune nimic despre ele, iar pe de altă parte tema propriu-zis probabilistă nu poate ocupa prea mult loc în acest mic volum fără a umbri ideile principale din această carte.

Tocmai în căutarea ieșirii din aceste dificultăți autorul a optat pentru ideea de a acorda expunerii, în această parte a cărții, o formă naivă, puerilă, care corespunde oglindirii necesarmente primitive a fondului problemei. Desigur că cititorul inițiat va răsfoi cu un justificat dispreț paginile următoare, pînă la pagina 129 inclusiv.

una, și alta, dar trebuie să aleg, fiindcă și așa sînt predispus la obezitate. De aceea m-am dus la ghicitoare, ea a aruncat o monedă, a căzut „capul", înseamnă că astăzi voi mînca prăjitură! Altfel nu m-aș fi putut decide niciodată ce să prefer și aș fi rămas cu totul fără desert, ceea ce ar fi fost foarte trist.

Noi: *Aveți pe de-a-ntregul dreptate. Și în lumea noastră s-a întîmplat o dată o chestie foarte tristă. Un oarecare Buridan avea un măgar. S-a întîmplat deci că plecînd pentru multă vreme de acasă, stăpînul i-a lăsat măgarului iubit două grămezi de fîn, dar, din nefericire, le-a făcut întru totul egale și le-a așezat la dreapta și la stînga măgarului la distanțe absolut egale. Ei bine, închipuiți-vă că măgarul lui Buridan a murit de foame, fiindcă nu s-a putut hotărî din care anume grămadă de fîn să mănînce.*

Epsilon: *În lumea noastră nu avem măgari, dar, chiar fără a cunoaște aceste animale, trebuie să presupun că ele nu posedă o inteligență deosebită. La noi este clar pentru toată lumea: dacă nu știi care din două lucruri să preferi, n-ai decît să dai cu banul și problema s-a rezolvat.*

Noi: *Bine, dar dacă aveți o preferință clară pentru un anumit lucru, îl alegeți neapărat?*

Epsilon: *Dar cum puteți să vă închipuiți una ca asta? Eu, de exemplu, prefer pește. Peștele îmi place de cinci ori mai mult decît carnea. Deci dacă aș proceda cum spuneți, ar însemna să nu mănînc niciodată carne. E de închipuit? Doar și carnea e bună la ceva. În lumea noastră ne place variația și cu toții alternăm toate în concordanță cu gusturile și preferințele noastre. În treacăt fie spus, aceste gusturi și preferințe noi le numim probabilități.*

Noi: *Înseamnă că cinci zile la rînd mîncați peşte şi în ziua a şasea carne?*

Epsilon: *Nu, nu m-ați înțeles deloc. La ce bun o asemenea regularitate plicticoasă? De fiecare dată mă duc la ghicitoare. Ea îmi cunoaşte problemele: alege un zar cu şase fețe şi-l aruncă. Dacă iese unu, înseamnă că a venit vremea să mănînc carne, dacă iese oricare alt număr înseamnă că o rog pe soție să-mi facă la prînz peşte. Totul este cît se poate de simplu.*

Noi: *Bine, dar ce faceți dacă preferința dv. pentru un lucru este nu de 5 ori, ci, de exemplu, de 36 de ori mai mare decît pentru altul? Sau dacă trebuie să alegeți nu între două lucruri sau treburi, ci între mai multe? Atunci cum procedați?*

Epsilon: *Ghicitoarea are în rezervă instrumente convenabile: noi le denumim mecanisme aleatorii. Pentru exemplul de care vorbiți este adecvată „ruleta" ; ghicitoarea o pune în funcțiune, apoi aruncă în ea o bilă ; ruleta se învîrteşte, se tot învîrteşte, iar cînd se opreşte, bila se află într-un anumit cîmp numerotat. Cîmpurile sînt în număr de 37. Dacă bila se opreşte la zero, înseamnă că trebuie să fac un anumit lucru, dacă însă se opreşte la oricare dintre celelalte 36 de numere, sînt obligat să fac altceva.*

Noi: *Mecanismul acesta este cunoscut şi în lumea noastră. Dar îl utilizează numai jucătorii inveterați, în casele de joc. Totuşi, este foarte incomod ca pentru fiecare ordine a preferințelor să ai nevoie de un instrument deosebit.*

Epsilon: *Desigur că da: tocmai de aceea şi alergăm la ghicitoare, care dispune de toate instrumentele necesare. E drept, unii oameni inteligenți spun că este suficient şi un ban. Dealtfel, iată-l pe Profesor; el ştie totul. Domnule Profesor, permiteți-mi să vă prezint pe acest străin,*

care se interesează mult de obiceiurile țării noastre. Explicați-le, vă rog, cum se poate ghici cu ajutorul unui singur ban ce ai de făcut cînd se ivesc multe treburi diferite?

Prof.: *Nu este prea complicat dar totuși va trebui să însemnăm, să scriem cîte ceva, și aceasta cere timp. Sînteți grăbit?*

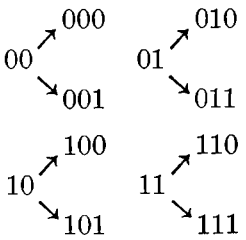
Noi: *Nu. Vom fi foarte atenți și vom nota totul în carnetul nostru.*

Prof.: *Atunci ascultați și priviți. Iată un ban, e absolut simetric, părțile sale pot fi deosebite una de cealaltă numai prin imaginile de pe ele. Ștrengarii numesc aceste fețe ale monedei „capul" și „pajura", noi însă le vom numerota pur și simplu. O față o vom numi 1, iar cealaltă 0. Dacă arunci banul, nu se poate ghici dinainte ce o să cadă, 0 sau 1. Ambele cazuri sînt egal posibile, egal preferate, egal probabile. Ați aflat, de bună seamă, că aruncarea cu banul este metoda clasică de a alege între două variante cu posibilități egale. Să încercăm însă a ne da seama cîte variante poate să aibă rezultatul aruncării unui ban de două ori. La prima încercare poate să fie 0 sau 1; aruncînd a doua oară sînt iarăși două posibilități 0 sau 1. Înseamnă că „istoricul" rezultatelor noastre poate avea 4, adică 2^2 variante. Să le scriem, folosind o notație simplă:*

00 01
10 11

În mod firesc avem două grupe: prima grupă se caracterizează prin faptul că la a doua încercare a ieșit; 0 a doua se caracterizează prin faptul că la a doua aruncare a ieșit 1. Dacă vom arunca cu banul încă o dată, atunci numărul variantelor posibile se va dubla iarăși,

ele vor fi în număr de 8 sau 2^3, deoarece fiecare din variantele vechi se bifurcă:



Lucrurile vor continua la fel odată cu fiecare aruncare. Prin urmare, numărul total al variantelor care pot să apară dacă aruncăm cu banul de N ori va fi 2^N; toate aceste variante sînt întru totul egal posibile. De aceea fiecăreia îi revine una și aceeași cotă de șanse de a apărea în fiecare caz concret în parte: $\frac{1}{2^N} = 2^{-N}$.

Această cotă o numim de obicei probabilitatea fiecărei variante concrete.

Astfel, devine clar cum se procedează pentru a alege între 2, 4, 8 și în genere 2^N acțiuni posibile, dacă nici uneia dintre ele nu i se poate acorda dinainte vreo preferință. Să presupunem, de exemplu, că este vorba de acțiuni: A_1, A_2, A_3 și A_4. Este suficient să convenim dinainte că operația A_1 va fi executată dacă de ambele dăți va cădea zero, A_2 va fi executată dacă la început va cădea 0 iar apoi 1, A_3 va fi executată dacă la început va cădea 1, iar apoi... Dealtfel, nu face să enumerăm, lucrurile sînt clare: fiecărei operații i se atribuie o anumită succesiune de zerouri și unități; operația al cărei șir „iese" va trebui să fie executată, realizarea șirului atrăgînd realizarea operației. Dar cum să facem dacă avem de ales între operații față de care avem anu-

mite preferințe? Firește că și aici principiul este limpede: operației pe care o preferăm îi vom atribui mai multe succesiuni „proprii" de zerouri și unități: îndată ce se va realiza una din ele, se va realiza și operația.

Dacă doriți, aș putea să vă vorbesc despre aceasta mai pe larg.

Noi: *(în unanimitate, deși obosiți): Da, da, firește.*

Prof.: *După cum probabil că ați înțeles, exprimăm preferințele prin numere exacte, probabilități. Dacă preferințele față de două posibilități A_1 și A_2 se află în același raport cu unu față de trei, atunci am putea spune că probabilitatea lui A_1 este 1, iar probabilitatea lui A_2 este 3. Este însă mai comod să nu avem de-a face cu numere oarecare; vom opera deci o schimbare a scării de proporții, de așa manieră încît* suma probabilităților tuturor cazurilor posibile, *adică probabilitatea alegerii a cel puțin unuia din ele, să* fie egală cu unitatea. *Este mai comod. Deci vom lua probabilitatea lui A_1 egală cu 1/4, iar probabilitatea lui A_2 egală cu 3/4. Atunci, după cum ați văzut, probabilitatea va corespunde cu cota șanselor fiecărei operații.*

Epsilon: *(care a ascultat și el cu mult interes relatarea Profesorului): Iată, de exemplu, eu vă spuneam cum îmi aleg desertul la prînz. Probabilitatea de a mînca ceva dulce este firește pentru mine egală cu unitatea, căci ceva dulce voi mînca neapărat. Iar întrucît șansele unei prăjituri sau a porției de cremă de ciocolată cu frișcă sînt egale, probabilitatea fiecăreia dintre ele este 1/2.*

Prof.: *Mulțumesc, Epsilon. Ai dat un exemplu foarte sugestiv. Îmi amintesc că în școală ai fost un elev capabil, dar apoi te-ai pasionat prea mult pentru aplicații, deși și asta e un lucru lăudabil. Permiteți-mi, totuși, să continui.*

Așadar, probabilitatea oricărei operații posibile, să zicem A, poate fi orice număr situat între zero și unitate; preferințe negative nu avem. Dar orice asemenea număr poate fi reprezentat întotdeauna cu aproximația dorită prin descompunere binară.

Noi: *Din păcate nu știm ce este descompunerea binară.*

Prof.: *Nu-i mare pagubă. Doar știți ce înseamnă o fracție zecimală. De exemplu, dacă în loc de 1/3 scrieți 0,333, aceasta înseamnă că ați scris 1/3 sub forma*

$$3 \cdot \frac{1}{10} + 3 \cdot \frac{1}{100} + 3 \cdot \frac{1}{1000},$$

sau

$$3 \cdot 10^{-1} + 3 \cdot 10^{-2} + 3 \cdot 10^{-3}$$

făcînd astfel o eroare mai mică de o mi..ie, adică mai mică decît 10^{-3}. În mod cu totul asemănător, orice număr între zero și unitate se poate reprezenta sub forma unei sume de puteri negative ale lui 2; tocmai asta se cheamă descompunere binară. De exemplu:

$$\frac{1}{3} = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^6 + \left(\frac{1}{2}\right)^8 + \left(\frac{1}{2}\right)^{10}$$

cu o eroare mai mică de 2^{-10}, ceea ce înseamnă aproximativ tot 10^{-3}. Această descompunere se poate nota și cu ajutorul unui rînd format din coeficienții puterilor consecutive

0,0101010101

și o asemenea reprezentare a lui 1/3 se numește fracție binară *sau* dual*ă. Este important faptul că aici se face uz numai de două cifre, 0 și 1: niciodată nu vom avea nevoie de altele, deoarece dacă întîmplător, din eroare, am rămas cu un*

termen de forma $5\left(\frac{1}{2}\right)^4$, *apoi un asemenea termen se poate totdeauna scrie sub forma*

$$(1+4)\left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^4 + 2^2\left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^4,$$

adică iarăși o sumă de puteri a lui 1/2.
Pentru ca să nu se întîmple asemenea erori, se poate face uz de o regulă simplă. Fiecare număr p (între 0 și 1) cu o aproximație pînă la 2^{-N} *se scrie sub forma fracției binare*

$$0,\ \alpha_1\ \alpha_2\ \alpha_3\ \ldots,\ \alpha_N$$

unde toți α_i *sînt fie zerouri, fie unități și se determină succesiv. Dacă* $p > 1/2$, *atunci* $\alpha_1 = 1$ *dacă p este mai mic, atunci* $\alpha_1 = 0$. *Dacă* $p_1 = p - \alpha_1 (1/2) \geqslant (1/2)^2$, atunci $\alpha_2 = 1$; în caz contrar $\alpha_2 = 0$. Dacă $p_2 = p_1 - \alpha_2 (1/2)^2 \geqslant \geqslant (1/2)^3$ *atunci* $\alpha_3 = 1$ *ș.a.m.d.*
Așadar, în această privință lucrurile sînt clare. Să revenim la aruncarea cu banul. Istoricul rezultatelor a N aruncări se reprezenta de asemenea printr-un șir de zerouri și unități; în total erau 2^N *variante posibile ale acestui istoric. Acum să împărțim toate variantele în* $N + 1$; *grupe,* B_1, B_2, ... B_{N+1} *astfel încît aceste grupe să nu aibă nici o variantă comună, adică după cum se spune,* să nu se intersecteze.
Grupa B_1 *va fi formată din toate variantele pentru care rezultatul primei aruncări era 1, iar celelalte oricum. Șansele unite ale tuturor variantelor care fac parte din* B_1 *sînt, evident, 1/2, adică probabilitatea ca prima dată să cadă tocmai 1. În grupa* B_2 *vor intra variantele la care primele două locuri sînt ocupate de 0 și 1, iar celelalte de oricare din aceste cifre. Este evident că nici una din variantele care aparțin lui* B_1 *nu se vor întîlni în* B_2. *Probabilitatea lui* B_2 *este egală cu probabilitatea apariției*

lui 01 în două aruncări succesive, deci $(1/2)^2$. *În continuare, să introducem în* B_3 *toate variantele care încep cu 001. Probabilitatea lor totală va fi* $(1/2)^3$; *etc. Continuăm astfel de* $N - 1$ *ori. Atunci vor rămîne numai două variante. Într-una vor apărea de* $N - 1$ *ori zerouri, iar la sfîrșit o unitate; în cealaltă vom avea numai zerouri. Aceste variante le vom denumi* B_N *și* B_{N+1}. *După cum am stabilit de la începutul convorbirii noastre, probabilitatea lor este* $(1/2)_N$. *Pentru ca totul să fie clar, să vă prezint tabelul descompunerii în grupe.*

Varianta	Tipul succesiunii	Probabilitatea
B_1	1 ...	1/2
B_2	01 ...	$(1/2)^2$
B_3	001 ...	$(1/2)^3$
.....	.....	.....
.....	.....	.....
B_N	000 01	$(1/2)_N$
B_{N+1}	000 00	$(1/2)_N$
	de $N - 1$ ori	

Firește, suma probabilităților tuturor grupelor va fi 1, întrucît fiecare dintre variantele posibile desigur că face parte dintr-o grupă sau alta. Iar acum, după ce am realizat această construcție, aproape că nu mai e nimic de făcut. Grupele noastre le putem asocia în grupe mai mari; de exemplu grupele B_2, B_4, B_6, B_8, B_{10} *le putem reuni într-o grupă B și atunci șansa ca să cadă cel puțin una dintre variantele care intră în B este*

$$\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^6 + \left(\frac{1}{2}\right)^8 + \left(\frac{1}{2}\right)^{10},$$

adică aproximativ 1/3.

Aș putea să dau formulele pentru regula generală, dar cred că veți înțelege mai lesne dacă mă voi mărgini să dau un exemplu.

Să presupunem așadar că vrem să lămurim necesitatea realizării operației B, pentru care preferința se exprimă prin numărul p = 0,719. Să construim descompunerea binară a acestui număr: cu aproximație de o miime, ea este următoarea:

$$\frac{1}{2}+\left(\frac{1}{2}\right)^3+\left(\frac{1}{2}\right)^4+\left(\frac{1}{2}\right)^5$$

sau, mai simplu, 0,10111.
Întrucît aproximația adoptată determină în descompunerea binară cinci cifre, să dăm de cinci ori cu banul. Dacă va avea loc cel puțin unul din următoarele evenimente:

B_1 *(la prima aruncare cade 1),*
B_3 *(în primele trei aruncări cad succesiv 001),*
B_4 *(la început cad succesiv 0001),*
B_5 *(la început cad succesiv 00001),*

atunci operația B trebuie efectuată. Vă rog să observați că pentru a lămuri acest lucru nu-i deloc necesar întotdeauna să aruncăm chiar de cinci ori. Ne putem edifica mult mai devreme. Iar pentru locuitorii țării noastre, care veșnic se grăbesc să facă o alegere aleatorie, acest amănunt este foarte important.

Din păcate mulți dintre ei uită repede cele ce au învățat la școală și sînt nevoiți, ca în vechime, să se adreseze ghicitoarei. De exemplu, d-ta, Epsilon, ai uitat totul.

Epsilon: *Nu, domnule Profesor. Am uitat cîte ceva, dar mi-am adus aminte totul îndată ce ați început să explicați; dealtfel, am și aplicat cele învățate. Pentru astă seară soția mi-a promis o plăcintă de mere sau un tort cu cremă. În genere vorbind, prefer de două ori mai mult tortul plăcintei. Dar am dat cu banul numai de două ori și a căzut „capul"; or, eu luasem „capul" drept unitate și întrebarea o pusesem cu privire*

la plăcintă; prin urmare n-am încotro, va trebui să merg acasă și să comand plăcinta. La revedere.

Zîmbind, profesorul i-a făcut un semn prietenesc cu mîna lui Epsilon, cel care se dă în vînt după dulciuri; am procedat și noi la fel și, obosiți cum eram de convorbire, voiam chiar să ne despărțim de Profesor, dar el ne-a pus obișnuita întrebare care se pune unor musafiri din străinătate: „Ei bine, cum vă place țara noastră?", astfel încît am fost nevoiți să recunoaștem cu sinceritate:

Noi: *Știți, explicațiile dv. ne-au ajutat în multe privințe, dar n-am înțeles totuși lucrul principal: cum se mai poate face în această țară vreo previziune, cum poate fi ea administrată, în definitiv cum se poate planifica producția de carne, pește, făină, zahăr pentru a hrăni milioane de asemenea mîncăcioși ca merituosul dumneavoastră elev domnul Epsilon?*

Prof.: *Într-adevăr, ați pus întrebarea principală, dar, din fericire, răspunsul este simplu. În țara noastră numai în aparență (probabil că și dv. ați avut această impresie) domnesc dezordinea și anarhia. În realitate, noi sîntem cea mai ordonată dintre toată țările civilizate, căci la noi domină implacabil o lege:* legea numerelor mari. *I se supune fiecare amator de dulciuri în parte, dacă face uz de foarte multe ori de același ansamblu de preferințe. I se supun și acțiunile globale ale milioanelor de locuitori, dacă au aceleași preferințe. Legea rezidă în faptul că atunci cînd alegerea la întîmplare se face în repetate rînduri și de fiecare dată cu unele și aceleași preferințe, adică cu unele și aceleași probabilități, atunci aproape sigur cota de cazuri în care se realizează alegerea unei acțiuni date este apropiată de probabilitatea ei. Dacă simpaticul meu Epsilon nu-și modifică preferințele, atunci e sigur că aproximativ o treime din zilele anului va mînca plăcintă cu*

mere, iar în celelalte două treimi iubitul său tort cu cremă. Dacă în oraș un milion de persoane care se dau în vînt după dulciuri au aceleași preferințe ca și Epsilon, atunci planificarea poate să le asigure satisfacerea acestor gusturi, organizînd aprovizionarea orașului cu materiile prime necesare pentru a prepara aproximativ 333 000 de plăcinte cu mere și 666 000 de torturi cu cremă.

Noi: *Dar domnul Epsilon nega orice regularitate a alegerii sale, el nu era de acord să mănînce carne după fiecare trei zile de hrană cu pește.*

Prof.: *Într-adevăr, nu există nici un fel de regularitate. Se poate întîmpla să fie și opt zile de pește la rînd, dar atunci aproape sigur ele vor fi compensate cîndva de două zile de carne. E drept, nu este exclusă catastrofa ca întregul milion de oameni ce se dau în vînt după dulciuri să dorească în aceeași zi tocmai plăcintă de mere. Dar pentru aceasta, după cum este lesne să ne convingem, șansa este doar de 1/3 la puterea un milion; mai bine zis, un asemenea eveniment, potrivit aceleiași legi a numerelor mari, poate avea loc în medie o singură dată în decurs de 300 de milioane de zile. Evident, nu este cazul ca planificatorii să ia în seamă o asemenea eventualitate.*

Noi *(cu mult interes): Stimate domnule Profesor, ni se pare că am înțeles acțiunea legii numerelor mari, am înțeles că lumea dv. nu poate fi judecată după prima impresie, ci trebuie să examinăm fie timp îndelungat comportarea cel puțin a unui locuitor, fie să însumăm impresiile oferite de mulți locuitori cu aceleași preferințe. Se pare însă că ne-am cam încurcat într-o problemă de logică. Ați vorbit despre variantele alegerii repetate între două evenimente, 0 și 1, fiecare avînd probabilitatea 1/2. Oare nu trebuia și atunci să se manifeste legea numerelor mari?*

Prof.: *Desigur că da. Pentru N mare, pentru un număr mare de aruncări cu banul cotele zeroului și a unității, „a capului" sau a „pajurii", sînt aproximativ egale.*

Noi: *Lucrul acesta ne este clar și din experiența noastră de viață. Dar dv. spuneați că fiecare variantă a rezultatului are una și aceeași probabilitate (1/2). Atunci cum poate să domine varianta în care numărul unităților și numărul zerourilor este același?*

Prof.: *Din păcate, ați comis obișnuita eroare logică. Numărul egal de zerouri și de unități apare nu într-o singură variantă, ci în foarte multe. De exemplu, în variantele atît de diferite*

01101010
10110001
01011001
10101001

precum și în multe alte variante posibile ale aruncării cu banul de opt ori. Demonstrația legii numerelor mari se bizuie tocmai pe faptul că aceste variante, avînd fiecare probabilitatea $(1/2)^8$, formează majoritatea. Nu vreau să vă expun această demonstrație, deoarece ar fi foarte obositor. Să examinăm însă toate variantele posibile ale aruncării cu banul în acest caz particular. Ele sînt în număr de $2^4 = 16$. Pentru a nu omite vreuna, să ne ținem de principiul bifurcării.

Numărul aruncărilor	Variantele rezultatelor							
1	0				1			
2	00		01		10		11	
3	000	001	010	011	100	101	110	111
4	0000 0001	0010 0011	0100 0101	010 0111	1000 1001	100 1011	1100 1101	1110 1111

Astfel, ne convingem că în șase cazuri avem două unități și două zerouri, în patru cazuri trei

unități și un zero, în alte patru cazuri trei zerouri și o unitate și numai în cîte un singur caz, după cum și era de așteptat, avem numai zerouri, respectiv numai unități. Firește, 4 este un număr prea mic pentru a verifica pe el legea numerelor mari, dar ne convingem că chiar și aici variantele cu număr egal de unități și zerouri sau cele apropiate de ele sînt dominante și, mai ales, că, cu cît abaterea de la ceea ce ne așteptăm conform legii este mai mare, cu atît asemenea abateri sînt în număr mai redus.*

În cazul unor probabilități diferite de 1/2, desigur că va apărea aceeași situație, numai că formal totul va părea mai complicat. Dacă acțiunii A îi corespunde probabilitatea p, iar celeilalte acțiuni $q = 1 - p$, atunci fiecare variantă concretă cu o succesiune dată de alegere a lui A și B va apărea în cazul unei repetări de N ori cu probabilitatea $p^n q^{N-n}$, unde n este numărul total al aparițiilor lui A, dar cota totală a tuturor variantelor care conțin aproximativ pN apariții a lui A se apropie rapid de unitate odată cu creșterea lui N. Pentru a înțelege viața în lumea noastră aleatoare, mai este important să vă lămuriți asupra numerelor aleatoare. Într-adevăr, alegerea fiecărei variante de acțiune este asociată unui anumit număr. Dacă amatorul de dulciuri alege plăcinta cu mere, atunci soția lui va fi nevoită să consume să zicem 50 g de unt, 500 g de făină, 1 kg de mere și cîte altceva — eu unul nu mă prea pricep în

---

* Cititorul care își amintește bine despre combinările, aranjamentele și permutările pe care le-a învățat la școală și-a dat firește seama că tocmai teoria acestora este necesară pentru calcularea variantelor în cazul general. Dacă îl interesează realmente știința aleatoriului, apoi el poate s-o cunoască din multe cărți bune. Cunoștințele școlare sînt cu totul suficiente pentru a citi [6], iar dacă cititorul cunoaște și elemente de matematică superioară, probabil că cel mai bun lucru care se poate recomanda este lectura cărților [17], [9], [21].

arta culinară. Dar este clar că dacă s-a ales tortul cu cremă, va fi nevoie de mult mai mult unt, să zicem 400 g, pe cînd de mere nu va fi nevoie deloc. Prin urmare, de alegerea pentru care a optat mîncăciosul depinde cantitatea întîmplătoare de unt, mere, făină care se consumă. În fond, tocmai aceste numere îi vor interesa pe planificatorii care determină cantitatea de alimente ce trebuie să intre în oraş. Şi aici le vine în ajutor legea numerelor mari: cantitatea de unt necesară pentru a satisface poftele unui milion de iubitori de dulciuri poate oscila la întîmplare de la 50 de tone (cînd toţi mănîncă plăcintă cu mere) pînă la 400 de tone (cînd toţi se aruncă asupra tortului cu cremă), dar legea numerelor mari zice că asemenea cazuri extreme aproape că nu se întîlnesc. Situaţia cea mai tipică este atunci cînd aproximativ o treime mănîncă plăcintă, iar două treimi tort. Deci aproape întotdeauna va fi nevoie de aproximativ*

$$1/3 \cdot 50 + 2/3 \cdot 400 = 317 \text{ tone de unt,}$$

$$1/3 \cdot 1000 + 2/3 \cdot 0 = 333 \text{ tone de mere etc.}$$

Estimaţia care se obţine prin însumarea produselor dintre fiecare variantă a numărului aleatoriu şi probabilitatea acestuia se numeşte valoarea aşteptată sau speranţa matematică *a numărului respectiv.*

Dar să amintim că aceste estimaţii sînt valabile numai atunci cînd acţionează legea numerelor mari. În cazul unui număr mic de repetări, consumul poate fi cu mult mai mare sau cu mult mai mic decît cel aşteptat. Firește, consumul depinde nu numai de numărul repetărilor, ci şi de felul în care variantele diferă una de alta.

---

* În lumea reală planificatorilor firește că le vine mai greu, deoarece aici scara preferinţelor este mult mai variată şi ei trebuie să ţină seama de acest lucru.

De exemplu, în ce privește untul și merele, variantele „plăcintă cu mere" și „tort cu cremă" diferă foarte mult, dar în ce privește făina oscilația va fi doar de 100 g, întrucît pentru tort va fi nevoie probabil de 400 g. Mai precis, consumul așteptat de făină pe zi și pe cap de amator de dulciuri va fi

$$500 \cdot 1/3 + 400 \cdot 2/3 = 433 \text{ g.}$$

Sînt posibile abateri de 66 g într-o parte și 33 în cealaltă parte. Dar plăcinte se coc mult mai rar. Prin urmare, este mai bine să apreciem abaterile medii ținînd seama de probabilitatea apariției lor. Pentru a nu se încurca cu semnul abaterii, de obicei se calculează pătratul ei și și se apreciază abaterea așteptată a unei asemenea mărimi ca

$$(66)^2\, 1/3 + (33)^2\, 2/3 = 2/9 \cdot 100^2,$$

denumită dispersia numărului *aleatoriu, adică în cazul nostru a consumului de făină. Și mai comod este să facem uz de rădăcina pătrată a dispersiei, denumită* abaterea medie pătratică*. *În cazul făinii, ea va fi de aproximativ 47 g. Legea numerelor mari se manifestă și prin faptul că abaterea medie pătratică a sumei rezultatelor multor alegeri consecutive crește ca și rădăcina numărului lor, iar suma însăși este, firește, proporțională cu acest număr, astfel încît creșterea absolută a abaterilor rămîne în urma creșterii consumului total și cota relativă descrește rapid. De exemplu, în acțiunea globală a unui milion de iubitori de dulciuri rolul relativ*

---

* Dacă sînt posibile n variante ale valorilor numărului aleatoriu, $x_1, x_2, \ldots, x_n$, avînd fiecare probabilitatea respectivă $p_1, p_2, \ldots, p_n$, speranța matematică va fi dată de

$$m_x = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n,$$

iar dispersia va fi:

$$D_x = p_1 (x_1 - m_x)^2 + p_2 (x_2 - m_x)^2 + \ldots p_n (x_n - m_x)^2.$$

al abaterilor se va reduce de $\sqrt{10^6}$, *adică de 1 000 de ori. Dacă oscilațiile așteptate în consumul de făină a unui individ se estimau la aproximativ 1/10, pentru un milion vom avea doar 1/10 000, adică 0,01%. Am ajuns astfel iarăși la concluzia că în planificarea pentru întregul oraș nu este cazul să se țină seama de ele. Dar se pare că v-am surmenat cu spusele mele. Trebuie să recunosc, așa sîntem noi, profesorii, tare ne place să explicăm, să predăm; ați fost niște victime.*

Noi: *(după ce ne-am adunat toate rămășițele de politețe): Vai, cum puteți să spuneți așa ceva?! Totul a fost extrem de interesant. Vă mulțumim foarte mult.*

Prof.: *Pentru puțin. Înainte de „adio" vreau doar să vă spun un lucru. Tot ce-am scris, numerele, denumirile, formulele, desigur că le puteți uita. Principalul este să nu uitați fondul, esența legii numerelor mari. Această lege este unicul fir călăuzitor în toate complexitățile lumii noastre, care realmente nu este simplă.*

Astfel s-a încheiat îndelunga noastră convorbire, iar odată cu ea și întreaga excursie în lumea aleatoriului.

Din păcate, modestele noastre cunoștințe despre alegerea la întîmplare vor trebui aplicate într-un domeniu nu chiar atît de dulce ca acela în care le aplica Epsilon, dar se spune că cu cît problema e mai dificilă cu atît mai dulce e succesul.

Așadar, să încercăm să descurcăm comportarea unuia din reprezentanții familiei de proceduri cunoscută sub denumirea de *căutarea la întîmplare.*

Pentru simplitate îl vom studia la început numai în problemele bidimensionale. Să privim obișnuitul desen „topografic" (fig. 8.2). Să presupunem că din punctul O se face un pas de lungime r într-o direcție aleasă la întîmplare. Linia de nivel pe care este situat punctul O împarte circumferința de rază r și cu centrul O în două arce.

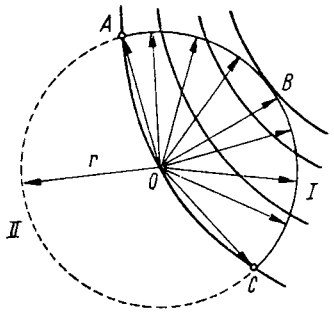


Fig. 8.2

Pe arcul I (să-l numim arcul reușitei) valorile funcției-scop sînt mai bune (nu mai proaste) decît în punctul inițial, iar pe arcul II lucrurile stau invers. Înseamnă că în cazurile cînd direcția aleasă face parte din sectorul $OABC$ pasul va fi mai mult sau mai puțin reușit. Dacă în cazul alegerii la întîmplare nici o direcție nu este preferată din capul locului, atunci este intuitiv clar că probabilitatea de a nimeri pe arcul reușitei este proporțională cu raportul dintre lungimea L a acestui arc și lungimea circumferinței.

Aceasta se poate dovedi, dacă determinăm care anume mecanism al alegerii la întîmplare este subînțeles în acest caz. Este clar că circumferința poate fi descompusă în 2^N arce, astfel încît pentru N suficient de mare aceste arce vor fi oricît de mici dorim; apoi se poate realiza alegerea unui arc aruncînd de N ori cu banul *. Atunci probabilitatea de a nimeri pe arcul reușitei va fi egală cu probabilitatea totală de a nimeri pe toate arcurile mici de lungime $2\pi r \cdot 2^{-N}$ care se așază pe ea. Dar numărul lor este

* În practică, atunci cînd se lucrează cu calculatoare se recurge pentru alegerea direcției la tabelele de numere aleatoare, tabele care se obțin la prelucrarea datelor experimentale cu privire la mecanismele aleatoare; este posibil de asemenea să se utilizeze chiar aceste mecanisme sau să se recurgă la formule speciale care imită în mod artificial caracterul aleatoriu al rezultatului (vezi cap. 10).

$\frac{L}{2\pi r \cdot 2^{-N}}$, adică probabilitatea totală va fi $L/2\pi r$, ceea ce trebuia dovedit. De obicei, o alegere de tipul celei efectuate de mecanismul nostru se numește *uniformă*.

Acum este clar că în cazul alegerii uniforme a direcției, probabilitatea reușitei (fie chiar relativă) este cu atît mai mare cu cît mai apropiată de o dreaptă este linia nivelului inițial în limitele pasului respectiv. Atunci arcul reușitei va coincide cu semicircumferința și probabilitatea reușitei va fi $1/2$. Aceasta este posibil în două cazuri: fie atunci cînd înclinarea este suficient de mare, asemănătoare cu panta acoperișului, fie atunci cînd lungimea pasului (adică raza cercului) este mică. Mai precis, înclinarea trebuie să fie suficient de mare în comparație cu lungimea pasului, ceea ce se întîmplă de obicei atunci cînd funcția-scop este suficient de regulată și cînd pînă la extremum mai e mult. Pe măsura apropierii de extremum, condițiile de obicei se înrăutățesc, arcul reușitei descrește și în vecinătatea nemijlocită a extremumului (la distanțe de ordinul lungimii pasului r) dispare cu totul, după cum se vede în figura 8.3. Așadar, probabilitatea

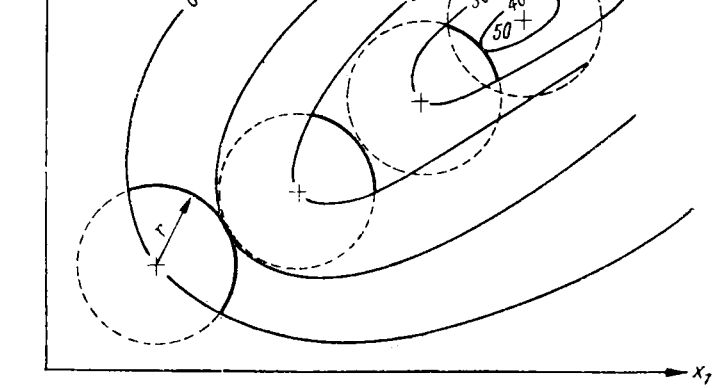


Fig. 8.3

reușitei în cazul alegerii uniforme a direcției dă rezultatele cele mai bune atunci cînd este utilizată departe de extremum. Să vedem acum cum se comportă căutarea la întîmplare în aceste condiții deosebit de favorabile.

Să figurăm tabloul liniilor de nivel la o scară mare, considerînd pentru simplitate că în limitele zonei demarcate înclinarea este constantă și egală cu k (fig. 8.4).

Să amintim că atunci cînd se face uz de schema gradientului pasul de lungime r va fi efectuat după direcția cea mai avantajoasă, perpendicular pe liniile de nivel; în felul acesta vom „cîștiga" o creștere a valorilor funcției egală cu kr. Însă pentru aceasta sînt necesare cel puțin trei măsurători.

Să examinăm acum acțiunea căutării la întîmplare. Aici, în caz de reușită, este necesară o singură măsurătoare; dar sînt posibile și nereușite. Mai întîi, să estimăm înaintarea la care ne așteptăm în cadrul unui anumit pas. Mecanismul aleatoriu îl folosim într-o formă foarte simplificată. Anume, vom descompune circumferința în trei părți egale prin punctele A, B, C, apoi vom face alegerea la întîmplare între cele trei variante ale direcțiilor (mijloacele arcelor I, II, III), considerînd că probabilitățile fiecărei variante în parte sînt egale cu 1/3. Atunci este clar că numai în cazul I vom obține succesul, creșterea valorilor funcției

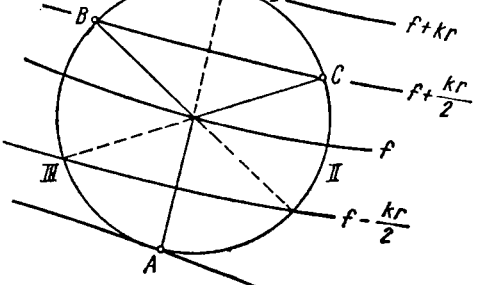


Fig. 8.4

fiind kr; în celelalte două cazuri vom fi nevoiți să ne înapoiem, adică creșterea va fi nulă. Așadar, valoarea așteptată este

$$\frac{1}{3} kr + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 0 = \frac{1}{3} kr,$$

adică pur și simplu o treime din creșterea pe care o obținem într-un singur pas folosind metoda gradientului.

Cititorul atent va fi observat desigur că am trișat puțin în cursul raționamentului, întrucît în prealabil am orientat direcția I după gradient, ceea ce firește că în virtutea caracterului aleatoriu (a descompunerii la întîmplare în trei arce) nu ar fi trebuit să aibă loc. Dacă însă am fi utilizat un mecanism mai rafinat de descompunere în 2^N arce cu N suficient de mare, atunci necesitatea orientării ar fi dispărut, iar calculele, firește mai anevoioase, ne-ar fi condus către un număr puțin deosebit de cel dat. Pentru N tinzînd către infinit pasul așteptat tinde către

$$\frac{1}{\Pi} kr \simeq \frac{1}{3{,}14} kr.$$

Concluzia calitativă fundamentală rămîne în vigoare: aplicînd schema căutării la întîmplare, creșterea medie a valorilor funcției-scop în cazul unui mare număr de pași este aproximativ de trei ori mai mică decît cea realizată cu același număr de pași dacă se face uz de schema gradientului. *Această concluzie rezultă direct din legea numerelor mari. Dar dacă ne amintim că în schema gradientului fiecare pas a consumat cel puțin de trei ori mai multe măsurători, rezultă că,* **dacă comparăm numărul măsurătorilor, căutarea la întîmplare este cel puțin tot atît de bună ca aceea efectuată după schema gradientului.**

Este deosebit de important că odată cu creșterea dimensiunilor problemei (dacă firește rămînem în condițiile tot atît de favorabile ale înclinării mari) încep să se manifeste avantajele căutării la întîmplare. *Pentru sferele multidimensionale (denumite în matematică hipersfere), calculul arată că numărul măsurătorilor necesare, k, în cazul*

căutării la întîmplare crește ca $\sqrt{n}$, *unde n este dimensiunea problemei, pe cînd cu metoda gradientului fiecare coordonată nouă pretinzînd o măsurătoare nouă rezultă că această creștere va fi proporțională cu n, adică mai mare (fig. 8.5).*

Așadar din comparația efectuată putem trage pe bună dreptate concluzia practică următoare: căutarea la întîmplare este indicată în problemele de mari dimensiuni, și anume ca un accelerator comod al startului. Prima porțiune a distanței, cînd pînă la extremum mai e departe, ea o parcurge rapid, fără multe măsurători de prisos; în a doua parte însă, se pare că e mai bine să înlocuim această metodă, încredințînd treaba mai fină a apropierii de extremum unei metode mai circumspecte, de exemplu aceleia a pantei celei mai rapide.

Dealtfel, noi nu am studiat decît comportarea unui singur reprezentant al familiei căutării la întîmplare, familie care, după cum am mai spus, este destul de mare și bogată. Între membrii săi există și unii mult mai circumspecți și mai delicați decît anarhistul pe care l-am descris. Desigur, cu toții **mențin principiul alegerii la întîmplare, la fiecare pas dat, a direcției de urmat, principiu valabil pentru întreaga familie, dar împrumută de la metodele**

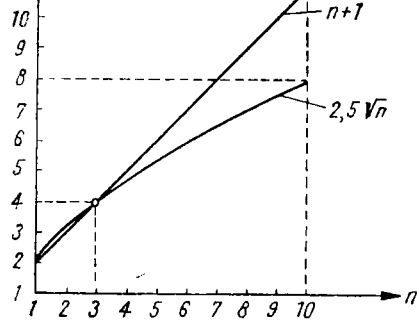


Fig. 8.5

care utilizează direcții convenabile ideea de a studia proprietățile funcției. *Într-adevăr, întrucît la fiecare pas, indiferent dacă este reușit sau nereușit, se execută măsurători ale valorilor funcției, aceste date pot fi utilizate pentru a prevedea comportarea funcției, pentru a învăța pe baza acestor cunoștințe.*

Să descriem principiul uneia dintre aceste **metode de învățare.** Să presupunem că direcția săriturii din punctul inițial se alege, ca și înainte, uniform aleatoriu, întrucît despre funcție nu știm încă nimic (fig. 8.6).

După prima încercare reușită, utilizînd ideea extrapolării, se poate presupune că această direcție va fi rezonabilă și ulterior, adică înaintînd după această direcție mai departe și efectuînd săritura următoare, vom nimeri iarăși într-un punct reușit. Desigur, nu putem crede pe deplin lucrul acesta; în orice caz, este indicat să restrîngem domeniul alegerii la întîmplare, reducîndu-l, de exemplu, de două ori; avem de ales unul dintre punctele situate pe semicircumferința orientată astfel încît mijlocul său să se afle pe direcția primei sărituri reușite. Apoi vom fixa două numere m_1 și m_2 (de exemplu, 2 și 6)

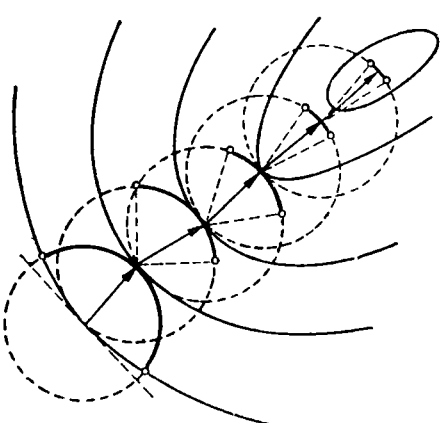


Fig. 8.6

și vom încerca iarăși să sărim într-un punct al semicircumferinței. Dacă reușita va veni după m_1 încercări sau după un număr de încercări mai mic decît m_1, atunci în punctul nou vom face alegerea în cadrul sfertului de circumferință orientat potrivit ultimei direcții reușite; dacă însă punctul reușit va pretinde un număr de încercări mai mare decît m_1, dar nu mai mare decît m_2, atunci și în punctul nou alegerea se va efectua în cadrul semicircumferinței; în sfîrșit, dacă numărul încercărilor nereușite va fi egal cu m_2, vom renunța de îndată la restrîngerea zonei de alegere și vom reveni la alegerea din întreaga circumferință, pînă vom ajunge la o variantă reușită (ceea ce va avea loc cu siguranță cîndva, în afara cazului cînd însuși punctul dat este extremal). Apoi în punctul nou vom încerca iarăși să restrîngem alegerea la semicircumferință. (Figura 8.6 reprezintă cazul cel mai reușit; fiecare pas a fost obținut în cel mult m_1 încercări.) O asemenea înaintare, cu restructurarea sectorului de căutare, seamănă cu căutarea care ar fi întreprinsă cu ajutorul unui felinar proiector, al cărui fascicul de lumină poate să fie făcut mai îngust și mai intens sau mai amplu și mai puțin intens. Sectorul restrîns sporește probabilitatea alegerii direcției corecte dacă această direcție se află în interiorul sectorului, pe cînd extinderea sectorului sporește această posibilitate.

Întrucît pe parcurs se poate estima și înclinarea medie între două puncte vecine, este rezonabil ca și lungimea pasului să varieze proporțional cu variația înclinării.

Există și multe alte metode cu învățare. Toate se întemeiază pe reglarea „sectorului vizionat" sau a lungimii pasului. Nu mă pot reține de a nu vorbi despre încă una, foarte simplă *. Din punctul inițial se execută un pas de lungime dată într-o direcție întîmplătoare, pînă se obține o reușită. Din noul punct prima încercare se efectuează exact în aceeași direcție și avînd aceeași mărime. Dacă

* Aici se descrie în mod simplificat una dintre variantele schemelor propuse de A.L. Gaiduk. Puteți face cunoștință mai îndeaproape cu aceste scheme interesante recurgînd la lucrarea [1].

prima încercare este de îndată reușită, atunci înaintăm în aceeași direcție, dar înmulțind mărimea cu un coeficient de creștere dat. Dacă însă prima încercare este nereușită, se face a doua în direcție opusă, lungimea pasului fiind împărțită la un număr dat. Dacă și această încercare este nereușită, trecem din nou la alegerea la întîmplare a direcției, reducînd lungimea pasului la fiecare nereușită. Atunci cînd în sfîrșit avem un succes, trecem în punctul nou și repetăm toate încercările în aceeași ordine, bineînțeles nu cu lungimea inițială a pasului, ci cu cea rezultată în procesul de învățare pe baza erorilor și reușitelor anterioare.

Așadar, am extins considerabil informațiile noastre cu privire la membrii tinerei și (oare de aceea?) întrucîtva ușuraticei familii de metode a căutării la întîmplare. Este clar că *în cazul unei instruiri suficiente și a unor condiții favorabile, metoda aceasta poate să concureze cu succes metodele din familia gradientului și din familia relaxării.*

În ce privește problema inițială formulată încă în cap. 7, dacă într-adevăr inventarea metodei pantei cele mai rapide este sau nu o realizare valoroasă, se poate răspunde categoric: da. Această apreciere este justă și pentru că *panta cea mai rapidă, precum și, în genere, schemele gradientului, posedă numeroase proprietăți utile pe care nu le au alte scheme* și pentru că, studiind atent această „invenție" și modificînd-o, se poate ajunge la o întreagă mulțime de alte invenții, de asemenea indubitabil utile; or, de bună seamă că aceasta este lucrul cu care se poate mîndri cel mai mult orice nouă invenție, mașină sau teorie!

Firește că cititorul insistent va lua iarăși neapărat cuvîntul și va întreba: „Bine, bine, dar ce metodă trebuie folosită? De analizat ați tot analizat, dar ce-a ieșit din asta?" Cititorului insistent i-am putea răspunde că încă n-a citit cartea pînă la capăt și că despre metodele căutării nu a aflat încă totul; nici pe departe. Dar, din păcate sau din fericire, nici în paginile finale ale acestei broșuri și nici în cărțile mai voluminoase el nu va putea găsi un răspuns direct la întrebarea sa. După cum vom mai repeta în numeroase rînduri, **totul depinde de alura funcției cu**

care avem de-a face în problemele concrete de căutare. *Dacă funcția-scop este de așa natură sau aproape de așa natură încît variațiile valorilor sale generate de variația unui argument dat se însumează pur și simplu cu variațiile generate de variațiile altor argumente, adică dacă* $f(x_1, x_2, \ldots, x_n)$ *diferă puțin de* $f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n)$, *atunci firește că nu se poate închipui ceva mai bun decît relaxarea după coordonate.*

Dacă variabilele se descompun într-un număr l de grupe independente, adică dacă funcția-scop diferă puțin de suma $f_1(x_1, x_2, \ldots, x_{m_1}) + f_2(x_{m_1+1}, \ldots, x_{m_2}) + \ldots + f_l(x_{m_l+1}, \ldots, x_n)$, atunci este rezonabil să efectuăm o relaxare suspendînd succesiv legăturile după fiecare grupă de variabile și căutînd extremumul pentru fiecare grupă în parte, de exemplu după schema gradientului (a se vedea mai în amănunt în cărțile [7], [11]).

Dacă toate variabilele sînt strîns legate și omogene din punct de vedere fizic, probabil că cel mai bine va funcționa schema pantei celei mai rapide.

Dacă variabilele nu seamănă deloc una cu alta și pot fi aduse la o formă omogenă, adimensională, numai prin alegerea scării, atunci nu sînt deloc proaste metodele pasului simplu, eventual și căutarea la întîmplare.

În plus, căutarea la întîmplare este aproape întotdeauna indicată ca un accelerator al startului, care permite lichidarea rapidă a marilor distanțări de extremum. Este clar un singur lucru: *pentru unele funcții-scop sînt mai bune unele metode, care nu sînt indicate pentru altele. Iar dacă alura funcției-scop, natura ei, ne este întru totul necunoscută, atunci nu putem face nimic altceva decît tot felul de încercări, organizînd „competiții" între metode pentru fiecare problemă concretă.* În acest caz, „în sport ca în sport", sau, parafrazînd, „în căutare ca în căutare": va învinge cel care este mai tare. Este limpede că **nu există campioni universali, în stare să obțină victoria în orice întrecere la care ar participa. E bine ca organizatorii competițiilor să țină seama de două principii practice: în primul rînd, dacă problema este asemănătoare cu unele deja rezolvate (de exemplu, se optimizează regimul unui cuptor în condiții cînd pentru cuptoare de același tip**

sau de tip asemănător căutarea optimului a fost efectuată), atunci este recomandabil să se utilizeze experiența cîștigată și să se înceapă prin a face uz de metodele încercate; în al doilea rînd, dacă încă n-au fost rezolvate nici un fel de probleme asemănătoare, este indicat să se înceapă cu metodele cele mai simple, de exemplu cu schemele relaxării sau cu schemele cu pas constant și numai ulterior, în caz de insucces, să se recurgă la metode mai complicate.

9
Diverse mici neplăceri

(Metode locale în condițiile mărginirii domeniului de căutare și a prezenței unor rîpe)

Tot mai adînc către abisuri conduc
cotiturile.
Valeri Briusov, *Glorie mulțimii*

La dreapta stîncă,
la stînga tot stîncă,
și între ele mărăcinișul pantelor line...
Rudyard Kipling, *Balada Răsăritului și Apusului*

Mîndri și cu capul sus pășeam din succes în succes, de la o metodă bună la alta nu mai puțin bună. O metodă era ceva mai inteligentă, alta era ceva mai rapidă, dar toate păreau să-și îndeplinească cinstit modesta lor sarcină: conduceau la extremum. E drept că se putea întîmpla ca acesta să fie doar un extremum local, dar, deocamdată, după cum ne înțeleseserăm, un asemenea succes ne satisfăcea întru totul.

Și totuși, pe parcurs apăreau din cînd în cînd diverse aluzii privitoare la tot felul de neplăceri de care s-ar putea ciocni schemele noastre favorite; de fiecare dată însă ne spuneam că e mai bine să amînăm discuția despre aceste lucruri neplăcute pînă vor veni timpuri mai bune sau, mai bine zis, mai proaste.

Acum însă, e timpul să nu mai trecem sub tăcere dificultățile și să declarăm deschisă o campanie care ar putea fi intitulată astfel: *„căutarea tuturor neplăcerilor ce se pot ivi în procesul căutării"*. Iar pentru ca această campanie să nu fie prea întristătoare, s-o asociem cu o alta: *„căutarea mijloacelor de luptă împotriva tuturor neplăcerilor ce se pot ivi în procesul căutării"* și să sperăm că nu vor avea loc nici un fel de accidente, eșecuri etc. Pentru

a ne întări încrederea în propriile forțe, să începem cu neplăcerile mărunte, să învățăm a lupta cu ele și treptat să trecem la altele mai mari.

Așadar, **neplăcerea nr. 1: mărginirea zonei căutării.** Să spunem dintr-o dată că pentru metodele căutării aceasta nu constituie un rău prea mare. Mai mult, nu putem nici măcar să fim siguri că mărginirea zonei ar fi un rău, deoarece, în general vorbind, cu cît zona este mai mică, cu cît sînt mai puține variantele discernabile, cu atît mai eficient se poate utiliza schema elementară a trierii cazurilor pe rețea. Dar pentru metodele locale cu iterare, de care ne-am ocupat atîta, prezența restricțiilor este firește o neplăcere.

Să amintim în primul rînd că printr-un extremum local înțelegem un punct mai bun decît toți vecinii săi, aceștia fiind aleși din domeniul admis, adică domeniul căutării. Punctul A din figura 9.1 nu are nici un fel de merite deosebite, maximul se află în B, dar dacă domeniul de căutare este restrîns, așa cum o arată frontiera punctată, atunci evident că punctul A va apărea ca un punct de maxim.

În acest paragraf ne vom ocupa numai de acele funcții-scop pentru care în domeniul căutat există un extremum (maxim sau minim) local unic. Cît de restrictivă este această ipoteză am arătat pe larg la sfîrșitul capitolului 4, dar chiar

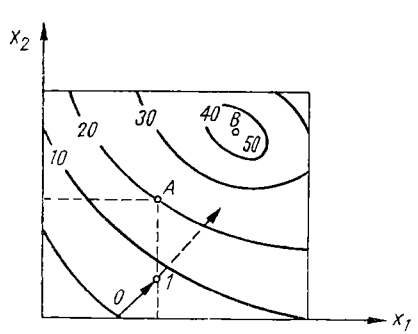


Fig. 9.1

și în aceste limite înguste metodele de iterare de care ne-am ocupat se lovesc de dificultăți reale.

Să începem cu schema pantei celei mai rapide. Să pornim dintr-un punct O situat în interiorul domeniului admis de frontiera punctată, așa cum se arată în figura 9.1. Să înaintăm după direcția gradientului (deocamdată lăsăm la o parte problema estimării acestei direcții). Această înaintare ne va conduce la punctul 1, care se află mai sus decît toate punctele de pe această direcție (cîtă vreme este vorba de punctele domeniului admis!). După oprire trebuie să determinăm direcția gradientului în punctul 1. Săgeata punctată din figura 9.1 arată rezultatul acestei determinări. După această săgeată ar trebui să ne mișcăm în conformitate cu metoda pantei celei mai rapide, ca, de altfel, și la toate celelalte scheme ale gradientului, dar n-avem cum: săgeata iese din cadrul domeniului admis. Așadar, iată că ne-am oprit; înseamnă că insuccesul este deplin, deoarece este sigur că 1 nu este punct extremal.

Astfel, necazul nr. 1 s-a manifestat din plin. Ce-i de făcut? Trebuie să căutăm în grabă *o posibilitate de a „lecui" schemele gradientului de prostul obicei de a stopa la frontierele zonei.* Din fericire, metodele de lecuire sînt numeroase, iar unele din ele extrem de simple.

Este evident că înaintînd pe direcția unei săgeți care formează cu gradientul un unghi ascuțit vom putea, de asemenea, să ajungem la puncte situate mai sus de cel inițial (fig. 9.2 *a*). După cum am mai spus în cap. 6, toate aceste săgeți determină direcții convenabile pentru căutare. Desigur, cu cît sînt mai apropiate de gradient, cu cît unghiul ascuțit este mai mic, cu atît mai rapid vom înainta pe versant. Dar ce-i de făcut dacă punctul se află pe frontieră și printre direcțiile posibile care nu ies din domeniul admis nu există nici una care să coincidă cu gradientul (fig. 9.2 *b*)?

Firește că cititorul va răspunde îndată: „Nu se poate merge după gradient? Ei și! Orice direcție e convenabilă, numai să fie posibilă, iar cea mai bună dintre ele este, probabil, cea mai apropiată de gradient, cea care formează

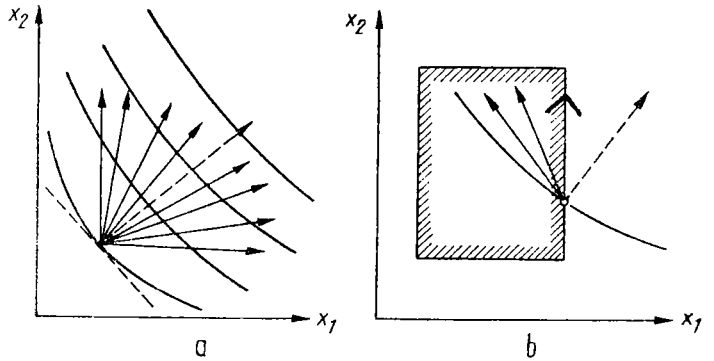


Fig. 9.2

cu el unghiul cel mai mic". Și cititorul va avea fără îndoială dreptate. Acest principiu simplu se află la baza majorității mijloacelor prin care metodele gradientului pot fi lecuite de deprinderea maladivă de a se opri aproape în fața fiecărui gard.

Atunci cînd în drumul său dă de un gard ce interzice înaintarea către țel, „omul care caută", dacă este rezonabil, va merge pur și simplu de-a lungul gardului, într-o direcție care să fie și convenabilă, și posibilă.

Tocmai restricția „gard" determină drumul nou, traiectoria nouă *. Aceasta trebuie urmată atîta vreme cît drumul ales este util, adică atîta vreme cît valorile funcției sînt crescătoare. Apoi se cuvine să ne oprim și să studiem din nou direcția gradientului. Această direcție poate să ducă în interiorul domeniului admis, în care caz o vom urma, procedînd ca de obicei; dar ea poate să și iasă din domeniul admis. Atunci va fi nevoie să facem o alegere între direcțiile convenabile și posibile, în caz că există asemenea direcții, și să înaintăm de-a lungul direcției alese (de preferință „lipindu-ne de gard"). Se poate însă întîmpla

* Deocamdată ne limităm la problemele bidimensionale, în cadrul cărora frontiera este unidimensională și deci determină traiectoria în mod unic.

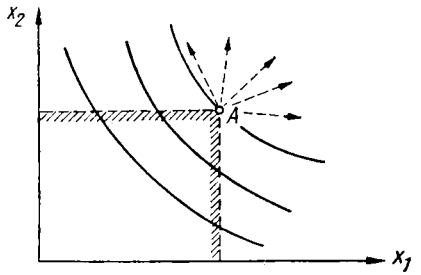


Fig. 9.3

ca în punctul în care ne-am oprit să nu avem direcții convenabile care să conducă în domeniul admis (punctul A pe fig. 9.3).

„Jalnică situație să n-ai încotro s-o apuci" — va spune cititorul pesimist. „Dimpotrivă, totul merge de minune!" — va obiecta cititorul optimist și va avea, desigur, dreptate: dacă s-a ivit o asemenea situație, într-adevăr n-ai încotro să mergi — dar nici nu ai de ce merge: tocmai punctul respectiv, punctul A, este extremumul căutat. Faptul că mulțimea tuturor direcțiilor convenabile care pornesc dintr-un punct nu conține nici o direcție posibilă este un criteriu clar și simplu, care arată că punctul respectiv este extremal (firește, după cum am mai spus: e vorba doar de extreme locale).

Așadar, metoda curativă căutată a fost găsită. Din păcate, ne-am mărginit doar la prezentarea unor figuri, dar n-am spus nimic despre regulile care guvernează schimbarea coordonatelor punctelor în procesul căutării. Din păcate, figurile, mai ales atunci cînd „mult" înseamnă mai mult decît „doi", nu sînt suficiente.

Să încercăm a depăși aceste neplăceri, la început în cazul bidimensional, și anume într-un exemplu foarte simplu *, cînd toate „gardurile" sînt drepte, mai exact

* Cele de mai jos constituie expunerea așa-numitei metode de proiectare a gradientului pe frontieră tocmai în cazul unei frontiere atît de elementare.

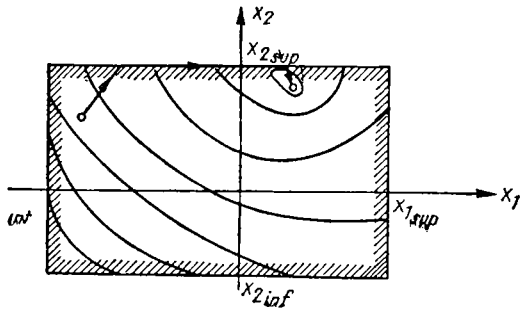


Fig. 9.4

alcătuite din sectoare drepte și orientate după axele de coordonate sau, mai simplu: cînd domeniul căutării este un dreptunghi (fig. 9.4). Un asemenea domeniu va apărea întotdeauna dacă pentru fiecare dintre variabile restricțiile se rezumă la existența unei margini superioare și a uneia inferioare:

$$x_{1\,inf} \leqslant x_1 \leqslant x_{1\,sup},$$

$$x_{2\,inf} \leqslant x_2 \leqslant x_{2\,sup}.$$

Atunci regulile vor fi foarte simple, întrucît pe frontieră una dintre variabile ia o valoare constantă (marginea superioară sau marginea inferioară) și numai a doua poate să varieze, astfel încît și formularea este elementară: dintr-un punct de coordonate $x_1^{(t)}$, $x_2^{(t)}$ trebuie să înaintăm după direcțiile determinate prin formulele:

$$x_1(\theta_1) = x_1^{(t)} + \theta_1 \left(\frac{\partial f}{\partial x_1}\right)_t,$$

$$x_2(\theta_2) = x_2^{(t)} + \theta_2 \left(\frac{\partial f}{\partial x_2}\right)_t$$

unde

$$\theta_1 = \begin{cases} 0, \text{ dacă } \left(\frac{\partial f}{\partial x_1}\right)_t > 0 \text{ și } x_1^{(t)} = x_{1\,sup} \text{ sau dacă} \\ \qquad \left(\frac{\partial f}{\partial x_1}\right)_t < 0 \text{ și } x_1^{(t)} = x_{1\,inf} \\ \theta, \text{ dacă condiția nu este îndeplinită;} \end{cases}$$

$$\theta_2 = \begin{cases} 0, \text{ dacă } \left(\frac{\partial f}{\partial x_2}\right)_t > 0 \text{ și } x_2^{(t)} = x_{2\,sup} \text{ sau} \\ \qquad \left(\frac{\partial f}{\partial x_2}\right)_t < 0 \text{ și } x_2^{(t)} = x_{2\,inf} \\ \theta, \text{ dacă condiția nu este îndeplinită.} \end{cases}$$

Dacă se va constata că $\theta_1 = 0$, atunci se va schimba numai x_2, iar x_1 va rămîne constant și egal cu cel de pe frontieră. Dacă $\theta_2 = 0$, atunci se va schimba numai x_1. Dacă θ_1 și θ_2 sînt amîndoi nenuli, înaintarea se execută după gradient, admițînd $\theta_1 = \theta_2 = \theta$. În toate aceste cazuri, înaintarea are loc pînă la maxim după direcția aleasă (numai în cadrul punctelor admise!). Dacă însă se va constata că $\theta_1 = 0$ și $\theta_2 = 0$, atunci, evident, punctul nou coincide cu cel vechi, mișcarea stopează și deci extremumul este atins. Întreaga schemă se transpune fără dificultate și în cazul multidimensional *: pentru orice coordonată se scriu aceleași formule ca acelea scrise mai sus pentru x_1 și x_2, schimbînd, firește, indicele. Este important de observat următorul lucru: în cazul tuturor coeficienților θ_1, θ_2 ,..., θ_n care nu se constată dintr-o dată a fi nuli, se admite că sînt egali între ei, și anume egali cu θ; de-a lungul direcției variază un singur parametru, θ, și toate coordonatele a căror variație este posibilă și de dorit variază simultan. Tocmai de aceea, dacă nici una dintre coordonatele punctului nu coincide cu valoarea ei

* Să amintim că e vorba de căutarea numai în asemenea domenii care sînt determinate prin existența unei margini inferioare și a unei margini superioare pentru fiecare variabilă. Dacă restricțiile au un caracter mai general, lucrurile se complică în mod substanțial, deși principiul direcțiilor convenabile posibile își menține importanța.

pe frontieră, ajungem la schema cunoscută a pantei celei mai rapide!

Acum putem spune definitiv că am găsit metoda de a ne salva de neplăcerea nr. 1. Este evident că această metodă rămîne valabilă și pentru alte scheme ale gradientului, care diferă numai prin alegerea parametrului θ, ca și pentru schemele pasului. Să vedem însă cum să lecuim metodele relaxării și, în genere, dacă este într-adevăr necesar să le lecuim. Mai întîi, e limpede că, chiar dacă este necesar, nu e întotdeauna. Într-adevăr, să ne amintim cum am tratat adineauri metodele gradientului care au făcut dificultăți pe frontierele dreptunghiului. Pur și simplu am constrîns una dintre coordonate să rămînă constantă, îngăduind schimbarea numai pentru cealaltă. Dar bine, tocmai asta este ideea relaxării: fixînd unele coordonate, să lăsăm să varieze celelalte! *Așadar metoda gradientului „lecuită" se comportă în parte ca o metodă de relaxare și tocmai aceasta îi permite să evite eșecul la frontieră.* Dar atunci este limpede că schemele relaxării n-au ce se teme de asemenea neplăceri. „Minunată concluzie, va spune cititorul atent, dar aceasta este valabil numai pentru dreptunghiuri sau paralelipipede avînd laturile (muchiile) paralele cu axele de coordonate; dar în cazul general?" Din păcate trebuie să recunoaștem că în cazul general treaba nu stă chiar atît de bine. În cazul general relaxarea, care permite înaintarea numai după direcțiile axelor de coordonate alese în prealabil, poate să apară ca o metodă prea îngustă. Din figura 9.5 se vede că în punctul A înaintarea după axa x_1 este în genere imposibilă (orice deplasare părăsește domeniul admis), iar deplasările posibile după axa x_2 conduc doar la descreșterea valorilor funcției-scop și, în consecință, sînt de asemenea nerealizabile în cazul relaxării după x_2. În punctul A schema relaxării va stopa, deși în limitele domeniului admis valoarea maximă corespunde punctului B.

Așadar neconcordanța orientării frontierelor cu direcțiile axelor apare ca o piatră de încercare pentru metoda relaxării. La prima vedere leacul pare clar: trebuie doar să rotim întrucîtva axele, să realizăm o transformare a coordonatelor,

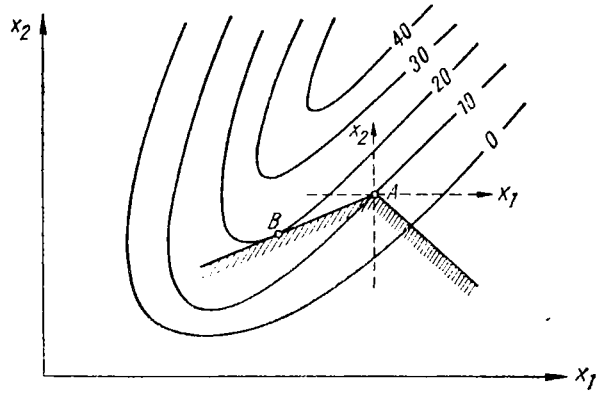


Fig. 9.5

de exemplu astfel încît direcția frontierei să coincidă cu direcția uneia dintre axe. Numai că nu este limpede cu ce frontieră să ne armonizăm atunci cînd punctul de oprire se află într-un colț; iar în cazul multidimensional (cînd „mult" este mai mult decît 2) asemenea frontiere pot fi destul de numeroase. În cazul general, relaxarea pură nu face față decît cu greu unor atare neplăceri. *Izbindu-se de ele, relaxarea trebuie să-și schimbe principiile și să facă apel la mijloace care-i sînt străine, recurgînd la alegerea liberă a direcției, care caracterizează metoda gradientului „vindecată", sau căutarea la întîmplare (după „vindecare" metoda gradientului este numită adesea „metoda direcțiilor posibile").*

Și acum, fiindcă a venit vorba despre căutarea la întîmplare: n-am pomenit nimic despre comportarea acestei metode în prezența unor frontiere. Aici situația este încurcată: uneori se comportă mai rău, alteori mai bine. Dacă în cazul alegerii la întîmplare a direcției se pot elimina de la început direcțiile imposibile (însuși mecanismul care generează „săriturile" întîmplătoare nu îngăduie apariția celor care depășesc limitele frontierei), atunci firește că lucrurile stau mai bine, deoarece șansele de a

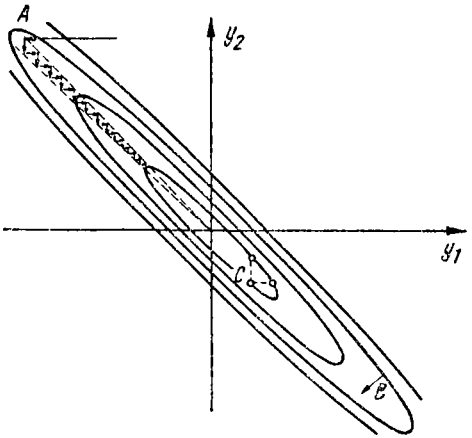


Fig. 9.6

alege ceea ce este de trebuință („arcul reușitei") sînt mai mari atunci cînd alegerea are loc pornind nu de la întreaga circumferință, ci doar dintr-o porțiune a ei. Dimpotrivă, lucrurile stau vădit mai prost dacă, așa cum se și întîmplă de obicei, mecanismul n-are habar de frontiere și generează orice direcții, în timp ce reușita reală nu constă în a nimeri pur și simplu „arcul reușitei", ci doar partea sa „posibilă", adică situată în interiorul domeniului admis. De aceea, majoritatea specialiștilor au adoptat următorul punct de vedere: **căutarea la întîmplare „nu simpatizează restricțiile", deși, într-un fel, le face față.**

Așadar am lămurit tot ce am putut în legătură cu neplăcerea nr. 1 și, în linii mari, am învățat să ne luptăm cu ea. Dar iată că apare un dușman nou și, probabil, mai perfid: **neplăcerea nr. 2, rîpele** *. Să privim la început figura 9.6. Curbele de nivel oferă imaginea topografică

* Cînd se caută maximul, atunci rolul rîpelor revine crestelor înguste. Evident, necazurile, ca și metodele de luptă, sînt aceleași.

a unei funcții-scop pe care o cunoaștem bine încă din cap. 7:

$$f(y_1, y_2) = \frac{1}{2}(a^2 + 1)(y_1^2 + y_2^2) - \frac{1}{2}(a^2 - 1)\, y_1 y_2.$$

Numai că aici mărimea a este egală cu 0,1, astfel încît elipsele liniilor de nivel constant sînt mult alungite și tabloul reprezintă nu o căldare, ci o rîpă. Perpendicular pe rîpă valorile funcției-scop se schimbă rapid, iar de-a lungul ei extrem de lent. Tocmai această împrejurare conduce la neplăceri serioase. Pentru început, să vedem cum îi face față schema relaxării. Foarte prost. Întrucît rîpa nu este orientată după vreuna din axele de coordonate, relaxarea din punctul A după y_1 sau după y_2 ne scoate aproape de îndată pe versantul rîpei și pașii devin foarte mici. Înaintarea merge încet, în zigzag (fig. 9.6).

După cum se vede, proasta comportare în cazul apariției unei rîpe este un viciu cert și incurabil al schemelor relaxării. *Dar poate că aici se comportă mai bine metodele gradientului, care, după cum se știe, au proprietatea alegerii direcției celei mai bune? Din păcate, și aici lucrurile stau prost.* În primul rînd această direcție este mai bună numai local. Direcția perpendiculară pe linia de nivel, pornind de exemplu din punctul B din aceeași figură 9.6, nu este de fel cea mai bună din punct de vedere global. În cazul pantei celei mai rapide este necesar să organizăm o căutare unidimensională foarte minuțioasă în această direcție, pentru a nu pierde fundul rîpei. Dacă însă se utilizează schema obișnuită a gradientului cu coeficient de proporționalitate fix, acest coeficient trebuie luat foarte mic, altfel pierderea este inevitabilă: pasul ne poate conduce pe versantul opus, într-un punct situat și mai sus decît cel inițial. Iar dacă coeficientul este mic, mișcarea va fi foarte lentă. În al doilea rînd — și lucrul acesta este, poate, mai important —, cînd dăm de rîpă trebuie să efectuăm cu multă atenție estimarea derivatelor. Priviți punctul C: din cauza cotiturii bruște a liniilor de nivel, valorile funcției în punctele de încercare deplasate cu Δ coincid cu valorile în punctul C. Estimațiile înclinării sînt aici

nule și căutarea pur și simplu se va opri fără a fi ajuns la extremum. De aceea, mărimea pasului de încercare trebuie să fie luată foarte mică, adecvată lățimii rîpei, iar aceasta, după cum ne vom convinge în cap. 10, mărește considerabil cerințele de precizie a măsurării valorilor funcției. În sfîrșit, mai există încă o împrejurare întristătoare: varianta prezentată în figura 9.6 nu e nici pe departe cea mai rea — există rîpe mult mai dăunătoare. Căci în natură rîpele și văile îndeobște nu sînt rectilinii: rîpele obișnuiesc să se încovoaie în fel și chip. Tot astfel și în căutare nu sîntem feriți de întîlnirea cu asemenea funcții ca cele a căror topografie este indicată în figura 9.7. Într-un asemenea caz, chiar dacă vom reuși întîmplător să nimerim exact în punctul de fund de pe însăși axa rîpei, primul pas pe care îl vom face ne va scoate de îndată spre un perete lateral, în timp ce în cazul unei axe rectilinii, cum se întîmplă în figura 9.6, un început reușit ne conduce de îndată la extremum.

Deci, cum să luptăm cu această grămadă de neplăceri cauzate de rîpe? Nu e greu de făcut față fiecăreia în parte. De exemplu, pentru a evita împotmolirea din cauza unei

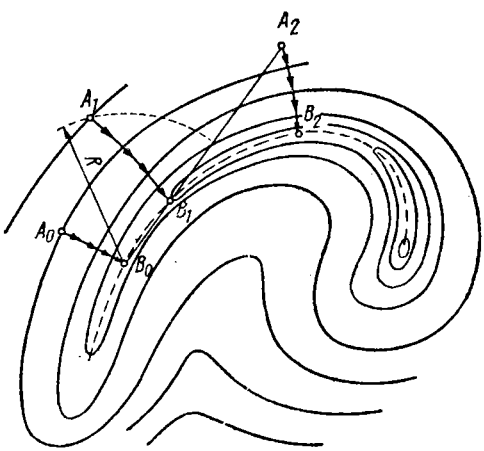


Fig. 9.7

estimații proaste a înclinării, putem nu numai să micșorăm mărimea pasului de încercare, dar și să recurgem la o altă schemă de așezare a punctelor de test: schema în cruce (fig. 9.8 *a*, *b*). În acest caz se execută pași de încercare de mărime $\pm \Delta$ în direcția fiecărei coordonate, iar estimația derivatei parțiale, de exemplu în raport cu x_1, se face astfel:

$$\left(\frac{\partial f}{\partial x_1}\right)_0 \simeq \frac{1}{2\Delta}[f(x_1^{(0)} + \Delta, x_2^{(0)}) - f(x_1^{(0)} - \Delta, x_2^{(0)})].$$

În cazul general o asemenea estimație „simetrică", bilaterală, ne dă o idee mai bună despre înclinarea în apropierea punctului *O*, dar, din păcate, ea nu ne salvează întotdeauna de împotmolirea departe de extremum dacă liniile de nivel se încovoaie foarte puternic (fig. 9.8 *a*). Totodată, pentru a o calcula sînt necesare aproape de două ori mai multe măsurători. În sfîrșit, pașii de încercare pot fi îndreptați nu în direcția axelor inițiale, ci după niște axe specifice fiecărei opriri, de exemplu se pot lua ca axe direcția care a condus în punctul respectiv și

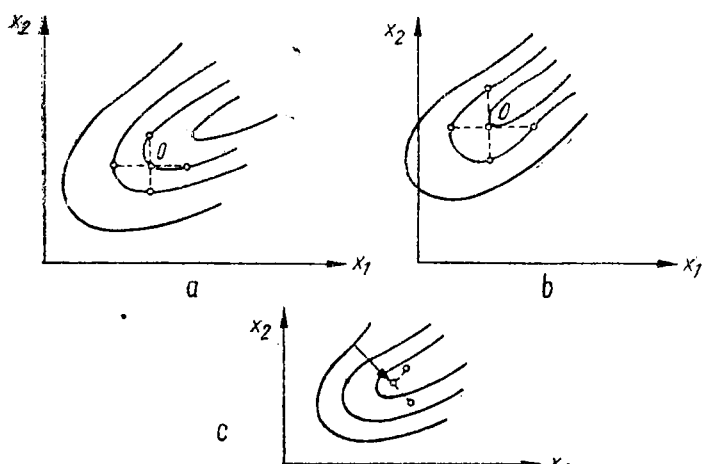


Fig. 9.8

cea perpendiculară pe ea (fig. 9.8 *c*). Deși un asemenea procedeu este util, calculul derivatelor este în felul acesta mult mai anevoios *.

Nici una din aceste măsuri nu ne ajută însă să luptăm împotriva celei mai mari neplăceri: tendința procedurii de căutare de a se ciocni de pereții rîpei ca un cățel orb, în loc să urmărească lin fundul ei ușor înclinat palpînd toate neregularitățile posibile.

E momentul să recunoaștem că nu avem la dispoziție o metodă care să ne garanteze izbăvirea reușită și deplină de toate necazurile legate de rîpe.

Există însă o metodă ingenioasă care, dacă este aplicată cu iscusință, promite succese remarcabile. Să-i expunem ideea principală.

Începem căutarea dintr-un punct arbitrar A_0 (fig. 9.7) și, aplicînd o schemă locală (de exemplu metoda clasică a gradientului cu coeficient de proporționalitate mic și constant), executăm cîteva iterări. Cît timp mișcarea are loc după versantul rîpei, înclinarea va fi suficient de mare, astfel încît chiar în cazul unor pași mici valorile funcției-scop vor varia în mod considerabil. Dar îndată ce coborîrea ne va duce la fundul rîpei, în apropierea axei sale, mișcarea se va încetini sau va stopa în mod automat și valorile funcției-scop aproape că vor înceta să se mai reducă. Dîndu-ne seama de aceasta, să ne oprim (în fig.9.7 este indicat punctul de oprire B_0). Desigur, s-ar putea întîmpla ca oprirea să fie determinată de un minim real; o asemenea ipoteză merită să fie verificată. Pentru aceasta vom face un mare pas la întîmplare, adică vom sări din punctul B_0 undeva pe circumferința de rază R, pasul fiind mult mai mare decît pașii anteriori. Să presupunem că am ajuns în punctul A_1. Acum vom încerca să coborîm din acest punct după gradient pînă la oprirea automată, iar dacă vom constata o încetinire sensibilă a înaintării ne vom opri noi înșine; ajungem astfel la punctul B_1. Dacă B_0 nu era un minim, ci doar se afla pe fundul rîpei, B_1 nu

* O idee cu privire la aceste calcule va fi oferită ceva mai departe, în cap. 14. Lucrarea [28], cap. 5, descrie mai amănunțit schemele de căutare cu rotirea axelor.

coincide cu B_0, ci se află de asemenea pe fundul rîpei. Acum avem la dispoziție două puncte care determină o direcție; putem realiza un pas mare pe această direcție („pas de rîpă", cum se numește el, într-un limbaj cam tîlhăresc). Nu pare indicat să palpăm toate punctele pe acest traseu (dacă rîpa este suficient de curbă, atunci direcția B_0B_1 ne va scoate aproape dintr-o dată pe versantul care se află mai sus de ambele puncte B_0 și B_1); este rezonabil să sărim pur și simplu dinspre B_0 înspre B_1 (dacă B_1 este mai jos) cu un pas proporțional cu înclinarea medie a fundului rîpei între B_0 și B_1, adică pur și simplu după schema gradientului. Să presupunem că sărind am ajuns în punctul A_2. Vom coborî iarăși cu prudență, ajungînd în punctul B_2 de pe fund, și măsurînd înclinarea între B_1 și B_2 vom face un pas de rîpă. (Este evident că schema permite buna urmărire a curburii axei rîpei).

Iată așadar cum, cu ajutorul „pasului la întîmplare", care provine din aceeași familie ca și „căutarea la întîmplare" (ajutor pe care nu l-am folosit decît o singură dată!), **am reușit să ieșim din tristul impas „rîpos".** *E drept că dacă dimensiunea problemei este mai mare decît doi, atunci lucrurile sînt mult mai complicate. În asemenea probleme ne putem lovi de un întreg „strat rîpos"; înaintînd de-a lungul acestuia valorile funcției-scop se schimbă foarte puțin, în timp ce la ieșire se schimbă brusc. Atunci trebuie să ne gîndim cum să alegem chiar în interiorul acestui strat o direcție rațională, iar pentru a aprecia „gradientul după strat" va fi nevoie de cel puțin trei puncte (în cazul tridimensional), adică vom avea nevoie de doi pași la întîmplare. Mai mult, stratul considerat poate avea propria sa rîpă, propria sa vale de-a lungul căreia valorile funcției-scop variază și mai mult.*

În felul acesta, în probleme multidimensionale poate apărea necesitatea de a utiliza o întreagă ierarhie, o scară de coborîri după gradient, de la cele inițiale cu pasul cel mai mic pînă la cele cu pasul cel mai mare.

Evident, nu putem ghici dinainte, fără a cunoaște funcția-scop, cîte asemenea nivele sînt necesare. Dar, după cum au indicat făuritorii „metodei rîpelor", I.M. Ghelfand și

M.L. Țetlin, **în fiecare problemă în parte intuiția fizică permite de obicei să ne dăm seama care variabile sînt neesențiale, adică determină o variație mare, ușor de descoperit, a valorilor funcției-scop și care anume sînt variabilele esențiale care determină o variație lentă, rîpoasă.** *Ei au reușit să dea o strălucită dovadă de intuiție fizică de acest fel, rezolvînd problemele de căutare foarte complicate care apar în analiza lauegramelor* *.

De aceea am și spus mai înainte că **noua metodă a rîpelor nu dă o rețetă standard pentru lupta cu rîpele, ci indică doar o idee fundamentală, bizuindu-se în rest pe experiența și iscusința calculatorului.**

* Ce înseamnă aceasta și ce legătură are cu căutarea se va arăta ulterior, în cap. 14.

10
Iluzii, peste tot iluzii...

(Erori întîmplătoare)

Viltoarea va să înghită
Luntre, luntraș fără grai.
H. Heine, *Lorelei* *

Pînă acum am avut de-a face numai cu mici neplăceri, necazurile mari abia ne așteaptă.

Așadar, să facem cunoștință: **necazul nr. 3, erorile măsurătorilor.** Pînă acum presupuneam în mod tacit că valorile funcției-scop pot fi măsurate în orice punct, și anume cu o precizie absolută; cu alte cuvinte, ceea ce s-a măsurat este așa și în realitate.

Dar oricine a avut cît de cît de-a face cu măsurătorile, fie și cu măsurarea propriei sale înălțimi sau greutăți, știe că măsurători infailibile nu există. O afirmație de tipul „Costel este mai înalt decît Petrică exact cu 3 cm" nu poate fi absolut adevărată niciodată. În primul rînd pentru că măsurarea înălțimii se face cu un dispozitiv a cărui precizie nu este mai mare de 0,5 cm și, în al doilea rînd, pentru că însăși înălțimea corpului nu este o mărime constantă, ci e întrucîtva schimbătoare, oscilînd chiar în cursul unei singure zile cu $\pm$ 1 cm. Să ne ocupăm în primul rînd și mai ales cu erorile de măsurare. Aceste erori sînt de două feluri. Primul gen: *erorile sistematice.* Divi-

* *Din lirica universală*, tălmăciri de L. Blaga, ESPLA, București, 1957 (*N.T.*).

ziunile centimetrice de pe măsurătorul înălțimii desigur că nu separă intervale exact egale cu centimetrul etalon, sutimea acelui metru confecționat din aliaj de platină și iridiu care este păstrat la Paris sub șapte peceți. La fabricarea unui „metru" obișnuit nu se folosesc instrumente de precizie, iar în timpul utilizării sale lemnul se mai usucă, vopseaua se întinde și „centimetrii" își schimbă lungimea. De aceea, în realitate, diviziunea 172 cm poate să însemne 171 cm și o asemenea eroare de 1 cm se va menține la toate măsurătorile. Desigur, o asemenea eroare nu e greu de înlăturat, dacă este cunoscută: de fiecare dată se poate scădea centimetrul de prisos. De altfel, pentru problemele noastre, problemele căutării, asemenea erori sistematice nu sînt o pagubă. Pe noi ne interesează de fapt numai diferențele de mărimi sau pur și simplu ce este mai mare și ce este mai mic; or, dacă ambele mărimi măsurate vor fi mărite sau micșorate cu același ε, diferența dintre ele n-are cum să se schimbe.

Așadar, dacă măsurătorile succesive ale valorilor funcției-scop se execută cu unul și același instrument, erorile sale sistematice nu vor influența desfășurarea procesului de căutare. Cu totul altă poveste sînt erorile de-al doilea gen: *erorile variabile*. Acestea apar pentru că atunci cînd se execută măsurătorile cu unul și același instrument, în realitate condițiile de măsurare se modifică întrucîtva, oricît ne-am strădui noi să le păstrăm neschimbate.

Iată-l pe Costel cum s-a așezat cuminte sub scîndura măsurătorului, silindu-se să-și așeze bine picioarele și capul, să-și aranjeze îmbelșugata și neastîmpărata chică. Binevoitorul doctor privește și spune: „Bravo, Costel, ai crescut binișor în timpul verii: 175 cm". Apoi, aceeași procedură se utilizează pentru Petrică. Plin de rîvnă cum este, Costel îl urmărește cu severitate pe Petre ca nu cumva acesta, doamne ferește, să se ridice întrucîtva în vîrful picioarelor. Se constată că Petrică are 174,5 cm. Costel nu mai poate de bucurie: doar e mai înalt cu o întreagă jumătate de centimetru! Dar bunul doctor, chiar dacă privește cu multă atenție prin ochelarii săi, ar fi putut spune tot atît de bine că înălțimea lui Costel este de 174,8 cm,

iar a lui Petrică de 174,7 cm, deoarece diviziunile de gradație sînt groase; atunci întreaga diferență s-ar reduce la o zecime de centimetru; iar dacă i-am măsura încă o dată, iar Petrică ar mișca un pic gîtul, n-ar fi mare lucru să-l „întreacă" pe Costel cu vreo treime de centimetru! Astfel încît Costel n-are cu ce să se mîndrească; dacă ar fi un băiat cinstit și cult, i-ar întinde lui Petre mîna și i-ar zice: „E grozav, Petrică! Sîntem deopotrivă; deosebirea aceea de jumătate de centimetru e un moft de-al erorilor de măsurătoare". Chestiunea care din doi este mai înalt, Petrică sau Costel, este în definitiv treaba lor personală; dar din păcate erori neconstante apar și la măsurători mai serioase.

Caracteristica cea mai importantă a calității unei turbine este randamentul, care se află raportînd puterea produsă la puterea consumată. Mii de proiectanți și oameni de știință își dirijează eforturile pentru a spori randamentul turbinelor măcar cu 1%. Într-adevăr, ce înseamnă 1% în plus pentru o instalație cu o putere de 500 000 kW? Aceasta înseamnă încă 5 000 de kilowați, o putere suficientă pentru luminarea unui întreg oraș de mărime mijlocie.

Realizarea unui randament maxim este o funcție-scop mai mult decît merituoasă. Dar una și aceeași turbină poate lucra cu randament diferit, după cum sînt condițiile de funcționare, presiunea apei, debitul de apă prin turbină. Pentru ca o turbină hidraulică să funcționeze în modul optim pentru fiecare ansamblu dat de condiții exterioare, se aranjează un dispozitiv special care permite ca paletele ei să fie reglabile astfel încît să imprime șuvoiului de apă direcția optimă, prin urmare și randamentul să fie cel mai înalt posibil în condițiile exterioare date *. Ar fi bine să putem măsura randamentul și să determinăm prin căutare poziția optimă a paletelor. Din păcate, nu știm să măsurăm randamentul mașinilor mari cu erori

* Asemenea dispozitive nu se utilizează la toate turbinele, ci doar la cele care funcționează în cazul unor baraje nu prea înalte, cînd condițiile de lucru pot să varieze mult.

mai mici de 1%, deci asemenea măsurători n-au la ce folosi*. Iar cauza imposibilității de a măsura exact randamentul rezidă în aceea că el depinde nemijlocit de debitul apei care trece prin turbină și este foarte complicat să se măsoare exact numărul metrilor cubi de apă care trece vijelios prin ea secundă cu secundă. Dacă viteza apei ar fi aceeași în toate punctele secțiunii transversale, n-ar fi cine știe ce greu s-o măsurăm și, înmulțind cu aria secțiunii, să găsim astfel debitul; dar șuvoiul pulsează, vitezele se schimbă de la punct la punct. Evident, secțiunea s-ar putea descompune în părți, măsurîndu-se viteza în fiecare porțiune dată, dar lucrul acesta este anevoios din punct de vedere tehnic și tot nu îngăduie deplina eliberare de erori, determinate de variația rapidă a vitezei de la un punct al secțiunii la altul.

Dar poate că exemplele noastre sînt nepotrivite: unul din ele prea copilăros, și pe deasupra referitor la un instrument de măsură deosebit de prost, iar celălalt mult prea grandios, deci indisolubil legat de mari dificultăți de măsurare?

Din păcate însă, aceste exemple sînt tipice. Se poate spune cu toată certitudinea că nu există instrumente de măsură lipsite de erori. **Fiecare măsurare este purtătoare de eroare, iar aceste erori variază de la o măsurare la alta;** *evident, noi putem să nu le observăm atunci cînd ne interesează numai estimațiile în mare, dar fără îndoială că ele sînt prezente.* „Dar cum așa, va spune cititorul neîncrezător, doar există instrumente ultraprecise".

Da, firește, instrumente **ultra***precise există, dar ele nu sînt* **absolut** *precise. Iar cauza acestei deficiențe nu este nici insuficienta ingeniozitate și inventivitate a inginerilor*

---

* Cititorul curios va întreba: atunci cum se poate vorbi despre un randament de 49% dacă el nu poate fi măsurat exact? Din păcate, măsurătoarea se face exclusiv pe cale indirectă, artificioasă, utilizînd mici turbine model, executate cît mai asemănător posibil cu surorile lor mai mari. Dar vorbind deschis, autorul n-a fost niciodată încredințat că randamentul măsurat pe un model cu o precizie mai mare decît 1% ar coincide cu randamentul turbinei reale.

proiectanți, nici insuficienta pricepere a muncitorilor executanți; cauza stă în însăși natura lucrurilor.

Aproape toate dispozitivele de precizie sînt electrice. Purtătorii electricității sînt electronii și ionii. Figurile din manualele elementare de fizică prezintă adesea rînduri strînse de electroni și ioni care merg de la un pol la altul. Din păcate însă, asemenea reprezentări nu prea sînt adevărate. În realitate, purtătorii electricității se află într-o continuă mișcare dezordonată, haotică, cu atît mai energică cu cît temperatura este mai înaltă. E drept că acțiunea cîmpului exterior îi constrînge pe purtătorii electricității să se deplaseze treptat într-o direcție preferată, dar dansul lor nu încetează niciînd. Mișcarea aceasta nu seamănă deloc cu o paradă militară: tendința își croiește drum prin haos, dar nu-l înlătură. Mișcarea haotică a purtătorilor creează și modificarea haotică a cîmpului electric cunoscută sub denumirea de „zgomot caloric". Aceasta este sursa oscilațiilor tensiunii fiecărui conductor, oscilații care îl caracterizează. Ele nu sînt mari (la temperatura camerei și la o rezistență de 1 megohm abaterea lor medie este doar de 0,01 volți), dar în cazul măsurătorilor exacte influența lor este vădită. Ce-i drept, temperatura poate fi redusă; la zero absolut (—273°C) „dansul" caloric încetează. Dar, în primul rînd este imposibil să atingem cu adevărat zero absolut, iar în al doilea rînd rămîn alte necazuri. Elementul principal al oricărei aparaturi de măsurare precisă îl constituie tuburile electronice, în care tensiunea pe grilă amplifică sau reduce curentul electronilor care merg dinspre catod la anod. Dar sarcina electronului este o mărime, ce-i drept, mică dar finită, deci intensitatea curentului va oscila din cauza discontinuității curentului. Iar această discontinuitate nu este regulată, ci haotică, întrucît electronii, deși părăsesc catodul și sînt purtați de cîmp, nu fac totuși aceasta în intervale de timp rigid determinate *. Astfel apare o altă sursă de erori, denumită „zgomot de fond".

---

* Autorul cere scuze cititorilor pentru toate inexactitățile cuprinse în aceste explicații; ele se datorează în principal scurtimii explicațiilor.

Orice creștere a preciziei măsurării este legată de reducerea lungimii, timpului, tensiunii și, prin urmare, ne apropie în mod implacabil de lumea microparticulelor; or, această lume are propria ei viață și toate eforturile noastre de a organiza condiții exterioare constante și stabile nu pot determina pe deplin mișcările locuitorilor ei. *Desigur că în măsurătorile care se referă la problemele optimizării, pe noi, care trăim în macrocosmos și tocmai aici ne formulăm obiectivele, ne interesează nu comportarea individuală a fiecărei particule, ci doar efectul total, cum ar fi tensiunea sau intensitatea curentului. Este însă important să înțelegem că acest efect are un caracter global, fiind rezultatul mediei activității diferitelor entități ale microcosmosului.*

Ei bine, e timpul să amintim că microcosmosul este realmente o lume ciudată. Deși, firește, locuitorii ei nu se adresează nici unei ghicitoare, sînt înzestrați ei înșiși cu un mecanism *întîmplător, aleatoriu,* care le determină acțiunile.

În sfîrșit, cuvîntul a fost rostit! Erorile de care ne lovim atunci cînd încercăm să sporim precizia măsurătorilor sînt manifestări ale „straniei" lumi aleatoare, lume ai cărei locuitori se supun legilor probabilității și, din fericire, principalei dintre aceste legi: **legea numerelor mari.** Din fericire; pentru că *pe noi, în macrocosmos, ne interesează numai efectul global al acțiunilor uriașului număr de locuitori ai microcosmosului, iar în procesul de însumare al acestor acțiuni rezultatele fiecăreia în parte sînt nivelate, mediate, iar abaterile relative de la ceea ce este mediu și așteptat devin cu atît mai mici cu cît numărul participanților este mai mare.*

Numai că legea numerelor mari nu este o lege rigidă, dictatorială, ea nu interzice apariția unor abateri, deși cu cît aceste abateri sînt mai mari cu atît ele apar mai rar.

Așadar, lumea aleatoare nu se ascunde doar în sălile tenebroase ale caselor de joc; ea este trup din trupul lumii reale. Tot ce este aparent constant în macrocosmosul în care trăim este doar o consecință a legii numerelor mari, care interzice manifestări frecvente ale unor abateri sensibile de la acest nivel constant. Dar dacă privim cu mai

multă atenție, vom observa la bază inconstanța în veșnică agitație, aleatoriul lipsit de orice stabilitate *.

Acum, dîndu-ne seama de universalitatea și inevitabilitatea erorilor întîmplătoare, e timpul să ne întoarcem la căutare, să vedem cum va reacționa în fața acestor erori „omul care caută"!

În fond, lucrul acesta era clar chiar din exemplul nostru „naiv": dacă Costel este într-adevăr de o înălțime apropiată de cea a lui Petrică, este greu să ne dăm seama care dintre ei este mai înalt și care mai scund. Toate rezultatele comparației „se înecă" în erorile întîmplătoare; dar tocmai compararea mărimilor apropiate este ocupația de bază a „omului care caută"!

Pentru a înțelege mai bine toate dificultățile care apar de aici, să examinăm mai în amănunt cea mai simplă problemă unidimensională a căutării, rezolvată, să zicem, după schema gradientului (cu parametru proporțional) și să ținem seama de cîteva formule simple. Pentru aprecierea înclinării în punctul inițial vom face, ca de obicei, două măsurători: în $x^{(0)}$ și în $x^{(0)} + \Delta$. Dacă $f(x^{(0)} + \Delta) > f(x^{(0)})$, ne deplasăm spre dreapta, în caz contrar, spre stînga. Tactica este cea obișnuită, dar, din păcate, nerea-

---

* Cititorul nu trebuie să rămînă cu impresia că aleatoriul se manifestă exclusiv cînd ne apropiem de microcosmos. Erorile în măsurarea vitezei curentului de apă erau legate de exemplu de fenomenul întru totul macroscopic, dar de natură aleatoare, al turbulenței. De altfel, și faimoasele erori în prognoza vremii depind de asemenea în mare măsură de dificultatea de a măsura turbulența din atmosferă.

Cititorul își amintește, desigur, că adesea valorile funcției-scop nu sînt rezultatele unor măsurători, ci numai ale unor calcule după formule exacte, dar foarte complicate.

Poate măcar aici nu e loc pentru erori întîmplătoare? Din păcate, nici aceasta nu corespunde realității. Calculul nu se face niciodată exact, totdeauna sîntem nevoiți să recurgem la rotunjiri, oprindu-ne, de exemplu, la cinci semne după virgulă sau la cinci cifre semnificative. Atît omul cît și orice calculator se comportă astfel. Mărimile de care se face abstracție sînt mici, dar dacă calculul cuprinde operații numeroase, erorile de rotunjire se pot acumula, rezultînd de aici mărimi sensibile de care trebuie să ținem seama. Erorile de rotunjire sînt la baza unor formule cu ajutorul cărora calculatoarele produc succesiuni de numere cu proprietăți asemănătoare numerelor aleatoare.

lizabilă aici, deoarece putem măsura nu ceea ce este în realitate, nu valorile însești ale funcției-scop, ci valori denaturate de erori întîmplătoare; în loc de $f(x^{(0)})$ găsim

$$f^*(x^{(0)}) = f(x^{(0)}) + z_1^{(0)},$$

unde $z_1^{(0)}$ este eroarea care are loc la prima măsurătoare, iar în loc de $f(x^{(0)} + \Delta)$ măsurăm

$$f^*(x^{(0)} + \Delta) = f(x^{(0)} + \Delta) + z_2^{(0)},$$

unde $z_2^{(0)}$ este eroarea care are loc la a doua măsurătoare.

Să presupunem că nici nu bănuim prezența unei erori și, comparînd pur și simplu rezultatele pe care le-am obținut, estimăm înclinarea:

$$S^*(x^{(0)}, \Delta) = \frac{1}{\Delta}[f^*(x^{(0)} + \Delta) - f^*(x^{(0)})] =$$

$$= \frac{1}{\Delta}[f(x^{(0)} + \Delta) - f(x^{(0)})] + \frac{z_2^{(0)} - z_1^{(0)}}{\Delta}.$$

În felul acesta în loc de

$$S(x^{(0)}, \Delta) = \frac{1}{\Delta}[f(x^{(0)} + \Delta) - f(x^{(0)})]$$

obținem o înclinare denaturată de prezența erorilor de măsurare:

$$S^*(x^{(0)}, \Delta) = S(x^{(0)}, \Delta) + \sigma^{(0)}, \text{ unde } \sigma^{(0)} = \frac{z_2^{(0)} - z_1^{(0)}}{\Delta}.$$

Adică: erorile conduc la faptul că S^* se deosebește de S printr-un termen aleatoriu σ, eroarea înclinării. Dacă încercăm să facem uz de schema căutării proporționale, atunci pentru a determina atît pasul cît și, mai ales, direcția lui, va trebui să facem uz tocmai de înclinarea denaturată S^*, întrucît alta, reală, nici nu cunoaștem.

Astfel, punctul $x^{(1)}$ va fi determinat după formula

$$x^{(1)} = x^{(0)} + \theta\, S^*(x^{(0)}, \Delta)$$

și poate figura oriunde: la dreapta sau la stînga punctului inițial, mai aproape sau mai departe de extremumul căutat; totul depinde acum de valorile erorii întîmplătoare din măsurători. Să vedem ce se întîmplă cu elementul principal: alegerea direcției. Aici lucrurile sînt foarte simple.

Dacă $|S| > |\sigma|$, eroarea nu influențează alegerea direcției. Dacă eroarea σ este de același semn ca și S, de asemenea nu există nici un efect dăunător. Dacă însă σ va fi mai mare în valoare absolută decît S și de semn opus, atunci totul se va orienta în cu totul altă direcție decît se cuvine.

Am stabilit așadar că erorile întîmplătoare pot determina un pas într-o direcție greșită. Iar dacă o împrejurare atît de defavorabilă ca aceasta se va repeta și la alți pași, ceea ce nu-i de loc exclus, atunci întregul proces al căutării se va orienta nu către extremum, ci în sens invers. Să te apuci cu mîinile de cap, nu alta!

E însă momentul să ne amintim din nou că în lumea aleatoare acționează legea numerelor mari, care ne salvează. Pe calea către extremum trebuie să facem mulți pași, și la însumarea efectului lor se va manifesta vădit direcția fundamentală, preferată. Într-adevăr, valorile înclinării denaturate oscilează în jurul valorii adevărate, iar valoarea adevărată determină tendința implacabilă a mișcării către extremum.

Acum este clar că erorile întîmplătoare pot să-l abată pe „omul care caută" de pe drumul cel bun, să-l silească din cînd în cînd să facă cîte un pas eronat sau chiar mai mulți pași eronați și totuși tendința sa de a ajunge la extremum, inclusă în valoarea medie, reală, a înclinării se va afirma în ciuda lor. *Ca întotdeauna în lumea aleatoriului, direcția principală se manifestă ca o tendință, ca o* **preferință** *și nicidecum ca o regulă rigidă.*

Sub acțiunea erorilor, „omul care caută" își pierde adesea rațiunea. Sorbind din otrava amețitoare a erorilor, el, deși aproximativ își dă seama de drumul „spre casă", spre extremum, este mereu ispitit să pășească în direcție contrară. Această comportare nu este însă un proces regu-

lat, întru totul *determinat*; ea devine o conduită supusă aleatoriului.

În lumea științifică așa și este denumită: *rătăcire aleatoare* (mersul la întîmplare); iar pentru a sublinia că *o atare rătăcire nu este deloc un nonsens, că aici își croiește drum o tendință preferată*, ea este comparată uneori cu mișcările unui înotător beat într-un canal cu apă curgătoare [1].

Așadar, erorile încetinesc înaintarea către extremum — aceasta este prima manifestare a neplăcerilor posibile —, dar totuși tendința trebuie să învingă și, mai devreme sau mai tîrziu, ea va conduce la extremum. Din păcate, aceasta nu este suficient. În lumea determinismului banal căutarea se va opri atunci cînd valoarea înclinării se anulează. În realitate, dacă după măsurătoarea în vecinătatea punctului $x^{(k)}$ am avut

$$f(x^{(k)} + \Delta) = f(x^{(k)}),$$

atunci însăși regula căutării arată că $x^{(k+1)} = x^{(k)}$. Am putea să repetăm iar și iar măsurătorile în vecinătate, dar toate ar da același rezultat și toate punctele ulterioare ar coincide cu $x^{(k)}$. Dar în lumea aleatoriului lucrurile nu stau astfel. Să presupunem că măsurătorile la pasul al t-lea arată că

$$f^*(x^{(t)} + \Delta) = f^*(x^{(t)}).$$

Atunci desigur că formula căutării va spune că trebuie să admitem egalitatea lui $x^{(t+1)}$ cu $x^{(t)}$. Dacă însă vom repeta măsurători în același loc, în virtutea aleatoriului ele pot da alte rezultate: de exemplu se poate întîmpla ca

$$f^*(x^{(t+1)} + \Delta) > f^*(x^{(t+1)}),$$

---

[1] Această comparație aparține renumitului savant sovietic L.O. Mandelștam, specialist în fizică teoretică. Împreună cu elevii săi, el a făcut foarte multe pentru înțelegerea fenomenelor reale ale lumii aleatoriului. Rolul rătăcirilor aleatoare în căutare a fost studiat pentru prima oară de A.A. Feldbaum. Odată cu lucrările sale, inginerii automaticieni au făcut realmente cunoștință cu remarcabilele posibilități ale căutării.

deși $x^{(t+1)} = x^{(t)}$, și punctul următor se va deplasa spre dreapta. De aici rezultă două concluzii: în primul rînd, dacă în apropierea unui punct înclinarea măsurată este nulă, aceasta nu înseamnă că punctul nu poate să fie foarte depărtat de extremum; în al doilea rînd, cînd ne aflăm în vecinătatea unui extremum autentic, erorile de măsurare pot „camufla" acest eveniment.

Să amintim că în vîrf, în punctul de extremum, înclinarea este nulă; nu există vreo tendință preferată de mișcare într-o direcție sau alta. Aceasta înseamnă că, venind la extremum de exemplu dinspre stînga, „omul care caută" are șanse egale sub influența vaporilor îmbătători ai erorilor întîmplătoare să meargă mai departe la dreapta sau să se întoarcă înapoi, dar în ambele cazuri el coboară versantul. Prin urmare, în pofida tendințelor care se manifestă atît la dreapta cît și la stînga și care în permanență îl împing către extremum, „omul care caută" este sortit să rătăcească veșnic în apropierea țelului căutat! Exact „în apropierea", deoarece legea numerelor mari nu-i permite să se îndepărteze mult de extremum, deși s-ar putea să nu-i interzică aceste rătăciri aleatoare.

Dacă graficul funcției-scop este simetric în raport cu extremumul, înseamnă că și forțele de atracție la dreapta și la stînga sînt egale, prin urmare în cadrul rătăcirilor nu se va ivi vreo preferință către o anumită direcție a îndepărtării, astfel încît, în medie, în cursul unor rătăciri îndelungate poziția omulețului nostru va coincide cu poziția extremumului. Aceasta ne și dă ceva speranțe în ce privește posibilitatea de a determina poziția extremumului; da, dar simetria nu-i deloc prezentă întotdeauna. Un versant poate să fie mai abrupt decît altul. Pe un versant mai abrupt tendința care împinge către extremum este mai puternică. De aceea poziția medie la rătăciri se deplasează în direcția versantului mai lin și nu va coincide cu extremumul, astfel încît speranțele noastre nu sînt întru totul justificate. Este foarte important să ne dăm seama cum depind abaterile de la extremum de caracteristicile principale ale procesului căutării, adică de pasul de încercare Δ și coeficientul de proporționalitate

0, precum și de caracteristicile erorilor și proprietăților funcției-scop. Dependența se poate stabili destul de lesne.

Cititorul amator de formule poate examina calculele care conduc către această dependență pentru cazul simplu cînd funcția-scop este descrisă de formula $f(x) = -1/2\, Lx^2$, unde $L > 0$. Înclinarea adevărată va fi $f'(x) = -Lx$. Maximul se află în punctul pentru care $f' = 0$, adică pentru $x = 0$. Regula căutării este următoarea:

$$x^{(t+1)} = x^{(t)} + \theta[-Lx^{(t)} + \sigma^{(t)}],$$

unde $\sigma^{(t)}$ reprezintă eroarea în măsurarea înclinării în cel de al t-lea punct. Fie că $x^{(0)}$ este luat arbitrar. Atunci, ținînd seama de regulă și aplicînd-o succesiv, obținem

$$x^{(1)} = x^{(0)}(1 - \theta L) + \theta\sigma^{(0)},$$

$$x^{(2)} = x^{(0)}(1 - \theta L)^2 + [(1 - \theta L)\,\theta\,\sigma^{(0)} + \theta\sigma^{(1)}],$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x^{(N)} = x^{(0)}(1 - \theta L)^N + [(1 - \theta L)^{N-1}\theta\sigma^{(0)} + \ldots + (1 - \theta L)\,\theta\sigma^{N-2} + \theta\sigma^{(N-1)}].$$

Valoarea coordonatei în pasul al N-lea și prin urmare și abaterea de la extremum, întrucît acesta se află în originea axelor de coordonate, este formată din doi termeni: unul depinzînd de punctul inițial, iar celălalt — de erorile întîmplătoare. Dacă luăm θ astfel încît $q = |1 - \theta L| < 1$, atunci odată cu creșterea lui N primul termen va tinde către zero (deoarece condiția „q mai mic decît 1" este tocmai condiția de convergență către extremum în absența erorilor întîmplătoare), dar necazul este că al doilea termen nu tinde către zero. E drept că valoarea sa așteptată este nulă, deoarece fiecare dintre erorile $\sigma^{(t)}$ este în medie nulă, dar sînt posibile abateri întîmplătoare. Așadar, după foarte mulți pași $x^{(N)}$ va rătăci realmente în apropierea extremumului, valoarea medie coincizînd cu acesta (aici funcția-scop inițială este simetrică).

Pentru un N foarte mare

$$x^{(N)} = \theta[\sigma^{(0)}q^{N-1} + \ldots + \sigma^{(N-2)}q + \sigma^{(N-1)}],$$

unde pentru simplitate am luat $1 - \theta L > 0$, se poate dovedi că dispersia sumei (sau a diferenței) numerelor aleatoare independente este egală cu suma dispersiilor (a pătratelor medii) pentru fiecare din ele; prin urmare, notînd cu D_σ una și aceeași dispersie a fiecăreia dintre erorile $\sigma^{(t)}$,

$$D\{x^{(N)}\} = \theta^2[q^{2(N-1)}D_\sigma + \ldots + q^2D_\sigma + D_\sigma].$$

De aici, conform regulii de însumare a progresiei geometrice infinite rezultă că pentru N tinzînd către infinit

$$D\{x^{(N)}\} = \theta^2 D_\sigma \frac{1}{1 - q^2} = \frac{qD_\sigma}{2L(1 - L\theta/2)}.$$

Dar $\sigma^{(N)} = \frac{1}{\Delta}(z_2^{(N)} - z_1^{(N)})$ deci

$$D_\sigma = \frac{1}{\Delta^2}(D_{z_2} - D_{z_1}).$$

Prin urmare,

$$D\{x^{(N)}\} = \frac{\theta D_\sigma}{\Delta^2 L(1 - L\theta/2)},$$

unde D_σ sînt dispersiile egale ale erorilor la fiecare măsurătoare.

Să amintim că Profesorul din lumea aleatoriului ne-a vorbit despre o măsură foarte simplă a abaterilor întîmplătoare: dispersia. Ei bine, *în cazul rătăcirilor aleatoare, dispersia abaterii de la extremum (pentru curbele simetrice) este cu atît mai mare: 1) cu cît este mai mare dispersia erorilor întîmplătoare la măsurători, ceea ce, firește, este întru totul banal; 2) cu cît este mai mare coeficientul de proporționalitate, deoarece cu cît „omul care caută" este mai mobil, mai grăbit, mai tumultuos, cu atît se poate îndepărta mai mult de extremum în cazul rătăcirilor; 3) cu cît este mai lină funcția-scop în apropierea extremumului, deoarece și în cazul îndepărtării de el înclinarea, prin urmare și tendința către extremum sînt mici și „se îneacă" în erorile întîmplătoare și, în sfîrșit, 4) dispersia abaterilor este cu atît mai mare cu cît este mai mic pasul de încercare Δ. Ceea ce, firește, este limpede dinainte, deoarece cu cît mai apropiate sînt punctele luate pentru comparație, cu cît valorile adevărate ale funcției-scop se deosebesc mai puțin în aceste puncte, cu atît mai mare este rolul relativ al erorilor aleatoare, cu atît mai multe sînt șansele de a greși cînd trebuie să determinăm ce este mai sus și ce este mai jos* *.

* În multe privințe efectul erorilor întîmplătoare seamănă cu efectul micilor neregularități, a micilor movile și gropi. În ambele cazuri creșterea lui Δ este utilă. Deosebirea rezidă doar în aceea că „neregularitățile aleatoare" nu sînt stabile, ci oscilează, amintind nu atît un drum prost, cît o mare agitată.

Așadar, tabloul calitativ al rătăcirilor aleatoare în apropierea extremumului devine pe deplin clar și dintr-un șir de considerente generale intuitive. Or, lucrul acesta este foarte important pentru organizarea căutării, întrucît de obicei nu se cunosc în prealabil în mod exact nici proprietățile concrete ale funcției, nici caracteristicile erorilor, astfel încît în organizarea căutării trebuie să ne călăuzim tocmai după tabloul calitativ.

Concluzia calitativă principală este că nu trebuie să executăm măsurători în puncte pe cît posibil mai apropiate străduindu-ne, cum am afirmat mai înainte, să determinăm cît mai bine, să localizăm poziția extremumului. După cum am stabilit acum, avînd în vedere erorile întîmplătoare, asemenea precizări devin fictive, deoarece abaterile întîmplătoare sporesc.

După ce am enumerat și am lămurit consecințele principale ale neplăcerii nr. 3 — eroarea măsurătorilor —, e timpul să ne apucăm și de sarcina dificilă a luptei împotriva lor. În această privință, de obicei sînt puse pe primul plan, în primul rînd, lupta împotriva inexactității în realizarea scopului final, iar în al doilea rînd problema încetinirii ritmului de înaintare către el.

Mijloacele principale de luptă sînt două; ambele se bizuie pe utilizarea rațională a legii numerelor mari. Dintre acestea, primul, cel clasic, este destul de rectiliniu. Întrucît sursa tuturor necazurilor sînt erorile provenite din măsurători, să ne luptăm nemijlocit cu ele.

În mare parte necazurile „omului care caută" erau urmarea încrederii exagerate în rezultatele unor măsurători executate o singură dată. Dar dacă am face măsurători de cîteva ori și am folosi pentru estimarea înclinării rezultatele *medii* ale măsurătorilor succesive? Dacă măsurătorile sînt în număr într-adevăr mare, atunci indubitabil se va resimți efectul legii numerelor mari: abaterile de la valoarea medie reală vor fi mici. Dar și în cazul unui număr nu prea mare de măsurători media lor va indica o abatere considerabil mai mică de la media reală. De exemplu, să presupunem că fiecare vecin este cercetat de două ori, astfel încît estimarea înclinării se efectuează

ținînd seama de erorile a patru măsurători. Atunci obținem:

$$S^* = \frac{1}{\Delta}\left[\frac{f(x+\Delta)+z_2+f(x+\Delta)+z_4}{2} - \frac{f(x)+z_1+f(x)+z_3}{2}\right]$$

adică

$$S^* = S + \frac{z_2 - z_1 + z_4 - z_3}{2\Delta}.$$

Pentru simplitate, să presupunem că la fiecare măsurătoare sînt echiprobabile erorile $+\varepsilon_0$ și $-\varepsilon_0$, ceea ce echivalează cu alegerea valorilor erorii de către un mecanism întîmplător de tipul aruncării cu banul. Prin urmare, va fi suficient să calculăm cazurile posibile a patru aruncări succesive, ceea ce am notat deja în blocnotesul nostru în cursul lecției domnului Profesor din lumea aleatoriului.

În șase cazuri va avea loc o compensare deplină,[1] în opt vom avea abateri de la medie cu $2\varepsilon_0$ și numai în celelalte două cu $4\varepsilon_0$. Astfel că în medie după toate cazurile posibile vom avea următoarea abatere așteptată a sumei $z_2 - z_1 + z_4 - z_3$ de la zero:

$$\sqrt{\frac{6}{16}\cdot 0 + \frac{8}{16}\cdot(2\varepsilon_0)^2 + \frac{2}{16}(4\varepsilon_0)^2} = \sqrt{4\varepsilon_0^2} = 2\varepsilon_0,$$

adică abaterea lui S^* de la S ar fi $\frac{2\varepsilon_0}{2\Delta} = \frac{\varepsilon_0}{\Delta}$. Dacă am calcula la fel rezultatele cercetărilor efectuate o singură dată, adică estimația înclinării după două măsurători, ar rezulta o abatere medie pătratică egală cu $\frac{\varepsilon_0\sqrt{2}}{\Delta}$. Prin urmare, dublarea numărului de măsurători a determinat reducerea abaterilor medii de $\sqrt{2}$ ori. Dar încă domnul Profesor ne prevenea că în cazul luării mediei pentru N mărimi, valorile medii vor descrește de $\sqrt{N}$ ori, așa că n-am făcut decît să ne convingem de valabilitatea acestei reguli în cazul unui exemplu simplu. Așadar, luînd media cîtorva măsurători efectuate în unul și același punct,

reducem în medie erorile întîmplătoare în estimarea înclinării, prin urmare și toate neplăcerile ulterioare. Dar aceasta nu se poate face decît cu măsurători repetate, deci și cu cheltuieli suplimentare; iar cîtă vreme se fac măsurători, noi stăm locului, neameliorînd nici cît de cît funcțiascop. În plus, oricît am recurge la medie, posibilitatea erorii nu este totuși exclusă. E drept că dacă vom lua media unui număr foarte mare de măsurători, această posibilitate va deveni foarte mică; nu ne putem însă îngădui să reflectăm infinit de mult înainte de a face pasul decisiv. Probabil că cititorul își amintește ce s-a întîmplat cu biata Elsa din povestea fraților Grimm, care, tot gîndindu-se la necazurile posibile, nu se putea în nici un fel decide să facă ceva.

Nedorind să repetăm povestea bietei Elsa, putem sau, pur și simplu, să punem cruce problemei neplăcerilor, recurgînd la media unui număr mic de măsurători în nădejdea că astfel șansele unui eșec nu vor fi mari, sau să născocim ceva cu totul nou.

Acest „ceva cu totul nou" a fost născocit relativ recent și a primit denumirea de *aproximare stohastică* sau metoda aproximării în cazul erorilor întîmplătoare *.

Ideea care se află la baza acestei metode nu este defel complicată: dacă nu-i admisibil să stăm locului, atunci hai să ne mișcăm. Toată problema e cum să ne mișcăm. Să ne amintim că în cursul analizei rătăcirilor aleatoare ne-am dat seama de faptul interesant și simplu că abaterile de la extremum depind de coeficientul θ: cu cît acest coeficient este mai mic, cu alte cuvinte cu cît căutarea este mai neîncrezătoare, mai lentă, cu atît mai mici vor fi abaterile de la extremum în regimul rătăcirilor care se statornicește după un mare număr de pași. De aici urmează o consecință elementară: să-l facem pe θ egal cu zero sau cel puțin foarte mic și vom scăpa de rătă-

* Metoda a fost propusă relativ recent de către oameni de știință americani, la început pentru căutarea rădăcinilor ecuațiilor (H. Robins și S. Monro), iar apoi și pentru căutarea extremumului (T. Kiefer și T. Wolfovitz).

ciri! „Dar asta este o prostie — va spune cititorul indignat — dacă θ va fi nul, atunci va înceta orice căutare și vom rămîne în punctul inițial!"

Ca întotdeauna, cititorul are dreptate. Și totuși, în această idee absurdă există un sîmbure de adevăr. Ea poate fi împăcată destul de simplu cu bunul simț: **să începem căutarea cu un coeficient θ suficient de mare, dar cu fiecare pas să-l facem tot mai mic și mai mic, făcîndu-l să tindă către zero treptat, doar odată cu creșterea numărului de pași către infinit.** Atunci și varza va fi întreagă și capra va fi sătulă: căutarea va avea loc, iar rătăcirile întîmplătoare se vor amortiza treptat.

Așadar, ne-am decis că este rezonabil să facem apel la următoarea schemă de căutare:

$$x^{(t+1)} = x^{(t)} + \theta^{(t)} S^*, \text{ unde } S^* = S^*(x^{(t)}, \Delta)$$

este estimarea înclinării în punctul $x^{(t)}$, conținînd erori întîmplătoare, după cum am mai notat; $\theta^{(t)}$ este coeficientul care descrește pe măsura creșterii numărului de ordine t al pasului și tinde către zero cînd t tinde către infinit.

Dar aici este important să nu sărim peste cal, să nu pretindem o descreștere prea rapidă a lui $\theta^{(t)}$, deoarece riscăm să nu mai ajungem niciodată la extremum, chiar dacă drumul nu ne va fi împiedicat de pașii eronați determinați de erorile întîmplătoare.

Pentru a ne convinge de aceasta, să considerăm cazul suficient de firesc cînd înclinarea (adevărată) nu crește pe măsura înaintării către extremum (punctul de pornire $x^{(0)}$ este la stînga maximului, unde înclinarea este pozitivă), adică

$$S(x^{(t)}) \leqslant S(x^{(0)}) = S_0.$$

Atunci în virtutea regulii noastre de bază avem:

$$x^{(t+1)} \leqslant x^{(t)} + \theta^{(t)} S_0.$$

Aplicînd succesiv această inegalitate, rezultă

$$x^{(1)} \leqslant x^{(0)} + \theta^{(0)} S_0,$$
$$x^{(2)} \leqslant x^{(1)} + \theta^{(1)} S_0 \leqslant x^{(0)} + \theta^{(0)} S_0 + \theta^{(1)} S_0,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x^{(N)} \leqslant x^{(0)} + [\theta^{(0)} + \theta^{(1)} + \ldots + \theta^{(N-1)}] S_0$$

Constatăm aşadar că deplasarea rezultată prin N paşi efectuaţi din punctul iniţial, $x^{(N)} - x^{(0)}$, nu depăşeşte mărimea

$$[\theta^{(0)} + \theta^{(1)} + \ldots + \theta^{(N-1)}]\, S_0,$$

determinată de către suma coeficienţilor lui S_0. Dar dacă suma acestor coeficienţi pentru un număr infinit de paşi nu va creşte nelimitat, înseamnă că nu ne vom putea deplasa decît cu o distanţă mărginită, egală cu produsul dintre această sumă şi S_0. Or, punctul iniţial se poate întîmpla să fie la mare distanţă de maximul căutat. Înseamnă că chiar cu un număr infinit de paşi, tot nu se poate ajunge la el.

Că, într-adevăr, un asemenea paradox este posibil în cazul unei descreşteri rapide a lui $\theta^{(t)}$, ne convinge un exemplu simplu. S-ar părea că nimic nu ne împiedică să micşorăm $\theta^{(t)}$ înmulţindu-l la fiecare pas dat cu un coeficient pozitiv mai mic decît unitatea, adică punînd $\theta^{(1)} = \beta\theta^{(0)}$, $\theta^{(2)} = \beta\theta^{(1)} = \beta^2\theta^{(0)}$ etc. Atunci, ţinînd seama de formula progresiei geometrice, nu e greu de văzut că

$$\theta^{(0)} + \theta^{(1)} + \ldots + \theta^{(N-1)} = \theta^{(0)} + \beta\theta^{(0)} + \ldots + \beta^{N-1}\theta^{(0)} =$$

$$= \theta^{(0)}\,\frac{1-\beta^N}{1-\beta}.$$

Dar pentru N tinzînd către infinit β^N tinde către zero. Prin urmare, înaintarea de la punctul iniţial este limitată de mărimea

$$\frac{1}{1-\beta}\,\theta^{(0)}S_0.$$

Aşadar în cazul reducerii coeficienţilor se recomandă prudenţă. Să încercăm să reducem $\theta^{(t)}$ după o regulă de asemenea simplă, dar mai lentă, luînd $\theta^{(t)} = \frac{\theta^{(0)}}{t+1}$, adică $\theta^{(1)} = 1/2\ \theta^{(0)}$, $\theta^{(2)} = 1/3\,\theta^{(0)}$ etc. Atunci suma

$$\theta^{(0)} + \theta^{(1)} + \theta^{(2)} + \ldots = \theta^{(0)}\left[1 + \frac{1}{2} + \frac{1}{3} + \ldots\right]$$

va creşte la infinit odată cu creşterea numărului termenilor * şi temerile că nu vom putea ajunge la maximum se dovedesc neîntemeiate.

---

* Această proprietate se demonstrează destul de uşor. Demonstraţia se întemeiază pe faptul că suma poate fi descompusă într-o infinitate de grupe de termeni (1/2, 1/3 + 1/4, 1/5 + 1/6 + 1/7 + 1/8,...), astfel încît în fiecare grupă în parte suma termenilor este mai mare sau cel puţin egală cu 1/2.

Putem ajunge la aceste scheme și pornind de la cu totul alte considerente. Să presupunem că ne-am decis să facem uz de schema clasică a gradientului cu coeficient constant $\theta = \theta^{(0)}$; atunci, apreciind înclinarea în punctul inițial, ne deplasăm către punctul nou după regula

$$x^{(1)} = x^{(0)} + \theta\, S^*(x^{(0)}, \Delta).$$

Vom estima iarăși înclinarea și, obținînd $S^*(x^{(1)}, \Delta)$, schema ne dă indicația de a înainta către $x^{(1)} + \theta S^*(x^{(1)}, \Delta)$, dar, amintindu-ne de pericolul erorilor întîmplătoare, ne decidem a nu acorda prea mare încredere acestei recomandări și, luînd media între punctul nou și punctul precedent $x^{(1)}$, vom trece în realitate la punctul

$$x^{(2)} = \frac{x^{(1)} + [x^{(1)} + \theta S^*(x^{(1)}, \Delta)]}{2} = x^{(1)} + \frac{\theta}{2} S^*(x^{(1)}, \Delta).$$

Aici vom măsura înclinarea $S^*(x^{(2)}, \Delta)$ și iarăși drept $x^{(3)}$ vom lua valoarea medie a valorii recomandate de către schema uzuală și $x^{(2)}$, dar de astă dată, dînd dovadă de o mai mare neîncredere față de recomandare, vom prefera valoarea veche de două ori mai mult decît pe cea nouă și vom lua:

$$x^{(3)} = \frac{2}{3} x^{(2)} + \frac{1}{3}[x^{(2)} + \theta S^*(x^{(2)}, \Delta)] = x^{(2)} +$$

$$+ \frac{\theta}{3} S^*(x^{(2)}, \Delta).$$

Este limpede că această schemă cu utilizarea mediei ponderate, care exprimă preferința crescîndă pentru vechi, conservatorismul crescînd, este echivalentă cu schema deja descrisă a pasului tot mai mic, schema aproximării stohastice. Dar acum am reușit să deslușim în această schemă și vechea și binecunoscuta idee a utilizării legii numerelor mari pe calea luării în considerare a unor medii ale rezultatelor măsurătorilor.

Este foarte important că această idee este întruchipată într-un mod principial nou: noi nu „meditäm" pur

și simplu la infinit asupra rezultatelor, încercînd să vedem sub micile valuri ale erorilor întîmplătoare terenul solid al adevărului, dar și îmbinăm aceste meditații cu acțiuni, cu deplasarea treptată spre extremum.

Cititorul este probabil lămurit că diferitelor alegeri ale caracterului descreșterii lui $\theta^{(t)}$ le corespunde un grad diferit de „spirit conservator" la alcătuirea mediilor ponderate dintre rezultatele vechi și recomandările noi, deoarece este evidentă egalitatea

$$x^{(t)} + \theta^{(t)} S^*(x^{(t)}, \Delta) = x^{(t)}\left(1 - \frac{\theta^{(t)}}{\theta^{(0)}}\right) + [x^{(t)} + \theta^{(0)} S^*(x^{(t)}, \Delta)]\frac{\theta^{(t)}}{\theta^{(0)}}.$$

Cu cît $\theta^{(t)}$ descrește mai repede, cu atît schema este mai „conservatoare" și cu atît mai mare este pericolul să nu mai ajungem la țintă. Dar descreșterea prea lentă a lui $\theta^{(t)}$ corespunde unei încrederi exagerat de mari în recomandările întrucîtva ușuratice (avînd în vedere condițiile grele ale prezenței erorilor) de a ne deplasa prea repede, recomandări care nu îngăduie să distingem cu ajutorul legii numerelor mari tendința principală atît timp cît ea se mai manifestă clar *. Căci în vecinătatea extremumului nu vom reuși să ne dăm seama de ea. Se constată că toate proprietățile utile necesare le posedă șirul dat de legea $\theta^{(t)} = \frac{\theta^{(0)}}{t+1}$. Aceste proprietăți pot fi menținute și în cazul unei descreșteri mai lente a coeficienților, prin urmare și în cazul unui ritm general de mișcare mai rapid.

Unul din procedeele de accelerare se bazează pe un principiu foarte simplu. Întrucît alegerea corectă a direcției este totuși dominantă, repetarea în permanență a

---

* O cercetare mai exactă arată că reducerea prea lentă a coeficienților nu asigură compensarea deplină a erorilor întîmplătoare. Este necesar să fie satisfăcută condiția ca suma pătratelor lui $\theta^{(t)}$, adică

$$[\theta^{(0)}]^2 + [\theta^{(1)}]^2 + \ldots$$

să nu crească nemărginit.

pașilor greșiți este aproape improbabilă. De aceea, însuși faptul repetării de cîteva ori la rînd a alegerii uneia și aceleași direcții este o confirmare indirectă a justeței ei. Orice schimbare a direcției este suspectă și impune prudență. În virtutea acestor considerente se propune ca lungimea pasului să fie micșorată numai atunci cînd se schimbă direcția, cu alte cuvinte se adoptă regula

$$\theta^{(t)} = \frac{1}{n(t)+1}\,\theta^{(0)},$$

unde $n(t)$ este numărul schimbărilor de direcție în t pași.

Să amintim că, dacă extremumul se află în interiorul domeniului căutării, atunci, în cazul unor erori întîmplătoare, se poate întîmpla să-l sărim, iar după trecerea peste extremum direcția care anterior a fost corectă devine falsă și își pierde superioritatea. De aceea, aproape sigur că va avea loc o schimbare a direcției în sens opus, prin urmare și o reducere a lungimii pasului; dar totodată acest lucru este și necesar: în vecinătatea extremumului este indicat să ne mișcăm pe cît posibil mai prudent, reducînd treptat lungimea pasului către zero, cu oprirea finală în extremum.

În afară de schema descrisă, sînt utile și scheme de aproximare stohastică provenind din schema gradientului cu pas constant (și nu proporțional) și funcționînd după regula

$$x^{(t+1)} = x^{(t)} + \theta^{(t)} \begin{cases} 1, \text{ dacă } f^*(x^{(t)} + \Delta) > f^*(x^{(t)}) \\ -1, \text{ dacă } f^*(x^{(t)} + \Delta) < f^*(x^{(t)}) \\ 0, \text{ dacă } f^*(x^{(t)} + \Delta) = f^*(x^{(t)}). \end{cases}$$

Această schemă funcționează foarte bine și face față cu succes erorilor întîmplătoare. În cadrul ei se poate face uz și de metoda accelerării pe care am descris-o mai sus.

Astfel, am ajuns la concluzia că aproximarea stohastică este remediul universal pentru maladia erorilor întîmplătoare.

Dar..., ca întotdeauna există un „dar" împotriva oricărei hegemonii absolute a unei singure metode.

În primul rînd, compensarea deplină a efectului erorilor cu ieșirea exactă în extremum în schemele aproximației stohastice se obține plătind un preț mare: crește numărul pașilor, deci și timpul total al căutării. Or, în practică compensarea deplină nu este necesară; în mod obișnuit este suficient să nimerim într-o mică vecinătate a extremumului, obținînd însă acest rezultat într-un număr acceptabil de pași. Astfel încît, cel puțin în cazurile unde nu se cere o precizie prea mare, media simplă va fi mai eficientă.

A doua împrejurare care viciază schema aproximării stohastice este de altă natură.

Funcțiile-scop care caracterizează calitatea funcționării multor mașini nu rămîn neschimbate în timp. Sub influența condițiilor exterioare, aspectul acestor funcții se modifică. Firește că schimbarea funcției poate conduce la deplasarea poziției extremumului, la un fel de derivă lentă a acestuia.

Posibilitatea derivei nu permite să reducem la zero, nici măcar treptat, viteza căutării: stoparea chiar în apropierea extremumului actual este periculoasă, deoarece ulterior, din cauza derivei, extremumul se va îndepărta încet dar implacabil și astfel mașina va funcționa într-un regim care nu va fi mai bun. Or, reducerea vitezei căutării pînă la zero este o caracteristică a aproximării stohastice. **Prin urmare, acolo unde este posibilă deriva, este inadmisibil să facem uz de această schemă.**

Totuși, în pofida tuturor acestor obiecții, schemele aproximării stohastice sînt indubitabil un remediu eficient care asigură deplină înlăturare a efectului dăunător al erorilor, dacă o asemenea înlăturare este în genere posibilă și dacă ne-am învățat să manipulăm acest remediu. Cititorul atent ar putea obiecta că întreaga expunere a fost consacrată exclusiv problemelor unidimensionale. Dar **nici tabloul calitativ al influenței erorilor, nici cele două**

metode de luptă împotriva lor pe care le-am descris nu comportă modificări de principiu cînd trecem la scheme cu gradient sau cu pas constant multidimensionale. Cum anume se pot preciza estimațiile derivatelor parțiale în cazul măsurătorilor suplimentare se va indica în cap. 14.

Așadar în ultimele două capitole am reușit să facem față — în linii mari — dificultăților ce apar la utilizarea schemelor cu iterare. Și totuși...

11
Cinci metode de a cuprinde necuprinsul

(Probleme multiextremale)

Nici telegramă nu-i
Și nici scrisoare,
Ci numai jocul
Întîmplării oarbe.
Iar de te duci la gară
La-ntîmplare,
Sosește cineva,
Nu te-ndoi.
Viteslav Nezval, *Gara Wilson*

Tot sărind prin mlaștină, broasca a ajuns pe un mușuroi ceva mai înalt și șezînd confortabil privea de aici, cu ochii săi miopi de broască, întreaga depresiune mlăștinoasă pe care o cunoștea atît de bine. O privea și se fudulea: și de ce nu, de vreme ce ea s-a înălțat atît de sus, în timp ce cunoscuții ei țopăie undeva, departe, jos.

Dar, dragă cititorule, oare nu ne-am înfumurat și noi cînd ne-am însușit metodele căutării locale? Într-adevăr, dacă extremumul local nu e unic, și întîmplător nu l-am nimerit pe cel global, nu este oare prea devreme ca să ne bucurăm? Nu va fi oare extremumul local pur și simplu un punct mai înalt într-o mlaștină?

Nu este greu de înțeles că dacă funcția-scop se poate comporta în mod cu totul arbitrar, atunci încercarea de a-i căuta extremumul este o încercare de „a cuprinde necuprinsul".

Să privim cîteva clipe înapoi și să ne reamintim istoria încercărilor noastre. Avansînd ipoteza regularității funcției, am presupus la început că dificultățile vor putea fi ocolite. Numărul variantelor care trebuiau comparate pentru a evalua extremumul cu orice precizie dorită, determinată în prealabil, a devenit finit. Dar, din păcate,

cînd am examinat mai îndeaproape problemele multidimensionale, și acest număr a devenit de necuprins.

Atunci am pus pur și simplu cruce problemelor generale ale căutării și ne-am ocupat în amănunt de probleme locale, obținînd succese considerabile. Poate că s-ar cuveni să punem aici punct, dar ceva ne împiedică să ne liniștim. Cît despre studiul metodelor locale, el a fost nu numai util în sine, dar a și îmbunătățit întreaga noastră înțelegere a posibilităților căutării, dezvăluindu-ne multe idei noi: studiul proprietăților funcției în cursul procesului de căutare, folosirea interpolării și extrapolării și, în fine, a prognozei ne-a introdus în „ciudata" lume a aleatoriului.

Să încercăm așadar să facem uz de aceste noi idei și pentru rezolvarea problemei generale, globale, fără însă a porni cu speranțe exagerat de mari, deoarece, indubitabil, cuvintele privitoare la „cuprinderea necuprinsului" sînt adevărate.

Posibilități de utilizat sînt suficient de multe, ele sînt și foarte variate, dar în majoritatea lor se face uz de unul și același procedeu, pe care-l cunoaștem din lupta împotriva rîpelor: introducerea pasului la întîmplare. Acolo, în combaterea rîpelor, îl foloseam doar în stadiul inițial al căutării; acum vom încerca să facem uz de el în repetate rînduri.

Pentru a nu ne rătăci în varietate, să așezăm toate schemele într-o anumită ordine.

a) **Alegerea cu totul întîmplătoare.** Pentru a prezenta această schemă vom proceda exact la fel ca la schemele locale. Să presupunem că am ales un punct arbitrar A_0, în care am măsurat sau am calculat valoarea funcției-scop. Să notăm această valoare cu f_0 și s-o reținem, după cum vom reține și coordonatele punctului. Am început ca de obicei; dar acum, pe neașteptate, vom trăda principiile după care ne-am călăuzit înainte: nici un fel de „palpare" a vecinătății, nici un fel de cercetare a ei! Punctul A_1 îl vom alege nu în vecinătatea lui A_0, ci iarăși cu totul arbitrar. Să măsurăm valoarea în A_1. Ea este f_1. Dacă este mai bună, atunci o reținem, iar valoarea anterioară

o eliminăm din memorie. Dacă este mai proastă, nu face să ne mai gîndim la A_1; să-l uităm! Apoi o nouă alegere arbitrară, o nouă măsurătoare cu reținerea acelui rezultat care este mai bun. E clar că alegînd de fiecare dată punctul în mod arbitrar, există și cîte o șansă să dăm chiar de punctul cel mai bun. „Dar dați-mi voie — se va înfierbînta cititorul indignat —, am ajuns din nou la triere!" Firește, la triere; dar la o triere întîmplătoare, deoarece alegerea arbitrară a fiecărui punct înseamnă că coordonatele sînt determinate de o alegere uniform întîmplătoare (a se vedea cap. 8) din toate valorile atinse în cadrul domeniului de căutare. Este însă rezonabil să ne întrebăm dacă este de vreun folos trecerea de la trierea pe rețea la trierea întîmplătoare. Din păcate, nu putem da un răspuns bine determinat. Și în cazul trierii pe rețea fixăm întrucîtva arbitrar punctul inițial și ordinea trierii, desfășurarea procesului. Alegerea nefericită a punctului inițial poate sorti eșecului întregul proces, deoarece în timpul mărginit de care dispunem vom reuși să studiem doar o mică porțiune a domeniului căutării. Or, aici, în cadrul trierii la întîmplare ne încercăm norocul nu o singură dată, ci la fiecare măsurătoare succesivă. Pe de altă parte, putem alege o rețea foarte grosolană și, după ce am triat toate valorile pe ea, putem obține, cunoscînd gradul variabilității posibile, o estimație clară a mărimii extremumului. De altfel și în cazul trierii la întîmplare se poate obține o asemenea estimație, deși ea nu este tot atît de exactă, ci face apel la cuvîntul „probabilitate".

Cred însă că **nu merită să organizăm o dispută serioasă între aceste scheme: ideile la care apelează ele sînt mult prea debile.** Să trecem mai bine la altă schemă.

b) **Alegerea întîmplătoare cu coborîri intermediare.** Să luăm din nou un punct la întîmplare, A_0, dar să nu începem cu exagerări, ci să organizăm un proces local de coborîre * începînd din acest punct și continuînd cobo-

* Unele considerente cu privire la alegerea tipului de căutare locală au fost exprimate în cap. 8. Este evident că raționamentele care au loc aici se referă la orice schemă, cu condiția ca într-un timp rezonabil ea să ne aducă în vecinătatea unui extremum local.

rîrea pînă la oprirea în punctul B_0, unul din extremele locale în care valoarea funcției să zicem că ar fi f_0; reținem această valoare și coordonatele lui B_0. Alegem acum, la întîmplare, un nou punct A_1, și de aici coborîm către B_1. Dacă valoarea f_1 corespunzătoare punctului B_1 este mai bună decît f_0, ștergem din memorie pe cea precedentă și reținem doar valoarea lui f_1 și coordonatele lui B_1; în caz contrar vom considera că alegerea întîmplătoare a lui A_1 este nereușită și vom uita rezultatele ei, reținînd în memorie valoarea anterioară f_0. Apoi iarăși vom organiza o alegere la întîmplare, iarăși o coborîre locală și așa mai departe. Valoarea stabilită la sfîrșitul fiecărei etape este reținută atunci cînd ea este mai bună decît cele precedente. (Evident că această schemă este foarte apropiată de „metoda rîpelor", așa cum a fost descrisă în cap. 9, dar aici punctele succesive din care se pornește coborîrea nu sînt legate cu nimic între ele.)

Iată o procedură cu adevărat monotonă: valorile reținute nu pot să se înrăutățească. Dacă funcția se comportă nu prea dezordonat, se poate presupune că după un număr suficient de mare de etape vom fi vizitat toate extremele locale și vom obține pe cel mai bun dintre ele. **În principiu, nici un număr finit de etape nu dă garanția deplină a succesului, dar timpul căutării este practic limitat și de obicei cercetătorii se declară satisfăcuți de un rezultat dacă 5 pînă la 10 încercări ulterioare nu aduc o îmbunătățire a lui.**

Cunoscutul savant norvegian R. Frisch, unul din pionierii aplicării matematicii în științele economice, denumește această metodă *„colaborarea cangurului cu șoarecele"*. Cangurul nu prea vede, dar poate să facă sărituri mari. În plus, după cum este bine cunoscut, are pe burtă un buzunar foarte comod. Așadar, cangurul, pornind prin munți și văi să caute o depresiune comodă, îl pune pe șoricel în buzunarul său. Face o săritură la întîmplare și îi dă drumul șoricelului din buzunar; acesta ajunge pînă la cea mai apropiată gropiță și îl cheamă pe cangur: ia vino să vezi, e mai bună gropița asta decît aia veche?

Dacă nu-ți place, pune-mă iarăși în buzunar și mai fă o săritură. Așa și tot așa continuă călătoria-căutare...

Însă *demersul șoricelului (căutarea locală) pretinde de fiecare dată cam prea mult timp*. Poate că ar merita să renunțăm în genere la această diferențiere obligatorie între pașii mici, cărora le cerem să fie regulați, și cei mari, cărora le permitem să fie întîmplători? Să examinăm o altă posibilitate.

c) **Alegerea întîmplătoare cu distanțe preferate.** După ce am măsurat în punctul inițial A_0 valoarea f_0, să pornim din acest punct cu un pas într-o direcție arbitrară aleasă uniform întîmplător. Lungimea pasului o alegem de asemenea la întîmplare, dar în această alegere la întîmplare a lungimii nu vom proceda cu indiferență, ci vom stabili o „scară a preferințelor": pașii mici îi vom face cu o probabilitate mai mare, pașii mari — cu o probabilitate mai mică. De exemplu, scara preferințelor poate fi construită astfel: un pas de lungime Δ se alege cu probabilitatea $1/2$, cel de lungime 2Δ este ales cu probabilitatea $(1/2)^2$ și în general pasul de lungimea $k\ \Delta$ este ales cu probabilitatea $(1/2)^k$. Pentru a putea nimeri în orice punct al domeniului vom conveni că cea mai mare lungime admisă a pasului $N\Delta$ nu este mai mică decît cea mai mare distanță pînă la frontiera domeniului căutării*.

Să folosim acum un mecanism întîmplător înzestrat cu o asemenea scară de preferințe. Lungimea pasului pe care o alege acest mecanism, asociată cu alegerea direcției pasului, va determina un punct B_0, unde efectuăm o nouă măsurătoare. Dacă valoarea găsită aici este mai bună decît cea inițială, considerăm pasul ca fiind reușit; această valoare se reține, iar B_0 este luat drept noul punct de sprijin A_1, punct din care pornim noul pas la întîmplare. Noua alegere a punctului este întîmplătoare, dar în vir-

* Probabilitatea pasului maxim $N\Delta$ trebuie s-o luăm ca fiind egală cu $(1/2)^{N-1}$, adică probabilitatea pasului al doilea după mărime, pasul de lungime $(N-1)\Delta$. Atunci probabilitatea totală a tuturor cazurilor va fi egală cu unitatea, iar mecanismul întîmplător care realizează o asemenea scară de preferințe este tocmai aruncarea cu banul de N ori (a se vedea cap. 8).

tutea scării preferințelor noastre tinde către vecinătatea lui B_0, care coincide cu A_1. Dacă însă valoarea B_0 este mai proastă decît cea din A_0, atunci pasul nou se face pornind din A_0; prin aceasta se evită tendința spre puncte nefericit alese.

Apoi, ca de obicei, procedura se repetă, reținîndu-se cea mai bună dintre valorile funcției-scop pe care le-am găsit. **Această schemă păstrează și meritele alegerii la întîmplare și tendința către trecerea în revistă a unei mici vecinătăți, tendință proprie metodelor locale.** În aparență, ea diferă prea puțin de schema locală a căutării la întîmplare pe care am studiat-o, dar un amănunt important — posibilitatea unor pași mari — îi dă un caracter principial nou, universal.

Într-adevăr, să presupunem că punctul A_0 este un extremum local diferit de cel global. Atunci pașii mici vor fi întotdeauna nefericiți, dar, după fiecare nereușită, revenim în punctul inițial și, întrucît probabilitatea unui pas către orice punct al domeniului căutării este diferită de zero, mai devreme sau mai tîrziu ne vom desprinde de atracția exercitată de extremumul local respectiv și vom nimeri într-o poziție mai bună. Admițînd atracția, preferința către rezultatul deja obținut, ne-am călăuzit după o idee „localistă" cu care ne-am mai întîlnit: mai binele se poate afla numai în apropierea binelui. În schimb, introducerea pașilor mari revoluționează această concepție întrucîtva conservatoare și legitimează căutarea mai binelui în condiții radical diferite de binele cunoscut, statornicit.

Tocmai această îmbinare a abordării evolutive (locale) cu abordarea revoluționară (globală) face ca metoda căutării la întîmplare să fie universală. Este evident că în principiu utilizarea acestei metode nu pretinde nici măcar continuitatea funcției-scop. Dar (ca întotdeauna, a venit timpul să spunem „dar") **tocmai această universalitate este o slăbiciune a metodei:** *metoda aproape că nu ține seama de particularitățile funcției date și utilizează într-o prea mică măsură experiența acumulată în cursul măsură-*

torilor care oferă cunoștințe cu privire la caracterul comportării funcției.

De aceea este important să se ia în considerare și metode care, păstrînd universalitatea, sînt totodată purtătoare a ideii de „învățare".

d) **Alegerea la întîmplare, cu distanță și direcție preferate.** *Toate metodele care intră în această grupă reprezintă în fond scheme de iterare a căutării la întîmplare, scheme pe care le-am examinat încă în cap. 8, dar,* **spre deosebire de variantele locale, ele admit posibilitatea de a păși din orice punct curent în orice alt punct al domeniului căutării.**

Să presupunem că în urma unui pas reușit am trecut din punctul A_{t-1} în punctul A_t. Acum trebuie să alegem punctul pentru un nou experiment. *Am examinat deja trei posibilități: a) noua măsurătoare se execută într-un punct ales cu totul întîmplător, independent de cel anterior; b) noua alegere la întîmplare este supusă unei reglementări a căutării locale; c) alegerea la întîmplare se face dintr-o dată, dar cu preferință pentru vecinătatea punctului anterior.*

Firește că este posibilă și abordarea d): alegerea la întîmplare se face cu preferință atît pentru vecinătatea punctului reușit anterior A_t, cît și pentru direcția care a condus către acest punct reușit, adică direcția dinspre A_{t-1} înspre A_t. Firește că se poate ține cont și de măsura reușitei anterioare, adică de mărimea creșterii valorii funcției-scop la trecerea de la A_{t-1} la A_t.

Să descriem mai amănunțit o variantă posibilă a unei asemenea luări în considerare a „preistoriei" căutării.

Să admitem că pasul din punctul A_t trebuie să aibă două componente: una reglementată, cealaltă întîmplătoare. Partea reglementată a pasului se face de-a lungul razei care continuă segmentul $A_{t-1}A_t$, lungimea sa fiind proporțională cu diferența dintre f_{t-1} și f_t. În schimb, partea întîmplătoare se construiește ca în schemele de care am vorbit mai sus la punctul c.

În felul acesta, centrul de atracție nu numai este punctul A_t, ci extremitatea porțiunii reglementate a pasului.

Schema poate fi desăvîrșită prin introducerea unei scări variabile în funcție de reușita sau nereușita în încercările anterioare. Principiul de manevrare a coeficienților de scară este cel firesc: dacă pasul anterior a fost reușit, atunci se poate amplifica influența porțiunii reglementate, lăsîndu-ne și mai mult „în voia" direcției reușite de care am dat; dacă însă pasul n-a fost reușit, atunci, dimpotrivă, va fi amplificată partea întîmplătoare.

Un al doilea procedeu de perfecționare a schemei constă în a construi partea reglementată utilizînd informații nu numai cu privire la ultima porțiune a traiectoriei căutării (direcția segmentului $A_{t-1}\,A_t$), dar și cu privire la „preistoria" mai veche. În acest caz direcția pasului reglementat se obține făcînd o medie ponderată a direcțiilor pașilor anteriori, ponderile asigurînd superioritatea relativă a „noului" față de „vechi".

În fond, fiecare dintre metodele enumerate comportă o mare varietate de scheme care diferă între ele prin alegerea legii de preferință probabilistă, prin modul de manevrare a coeficienților din porțiunea reglementată și porțiunea întîmplătoare, prin ritmul de uitare a trecutului.

Diversitatea posibilităților asigură o suplețe mai mare a metodei, care în prezent pare a fi cea mai promițătoare pentru rezolvarea problemelor căutării extremumului global. Este evident totuși că adesea ea constituie un remediu mult prea scump la relele pricinuite de miopie.

Să ne închipuim următoarea situație: căutarea ne-a dus în punctul A_t, care se află nu departe de extremumul global E (fig. 11.1). Porțiunea reglementată ne va conduce în punctul A_{t+1}. Pasul următor va fi efectuat într-o direcție arbitrară cu atracție spre A_{t+1}. Probabilitatea unei reușite în acest caz nu este mare: trebuie doar să nimerim în porțiunea de plan mărginită de linia de nivel pe care se află punctul inițial A_t. Însă în cazul unei nereușite partea reglementată se va reduce, iar partea întîmplătoare, după cum ne-am înțeles, va crește; mai precis: va crește probabilitatea de a nimeri într-un punct mai îndepărtat de centrul de atracție. Or, în felul acesta, probabilitatea reușitei nu se va mări, ci, mai curînd, se va reduce. Pe

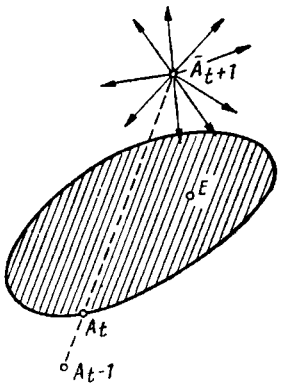


Fig. 11.1

de altă parte, orice procedură întru totul reglementată, de felul celei a gradientului, ne-ar permite în acest caz să ajungem foarte repede la scop. Să amintim că în genere metodele de căutare nu se prea remarcă prin cine știe ce merite în ultima fază a căutării atunci cînd ajungi în imediata apropiere a extremumului.

Raționamentele și estimațiile noastre se refereau la cazul cînd extremumul în apropierea căruia ne aflam era într-adevăr global și nu local. **Acum este clar că introducerea unui element întîmplător este plata pentru caracterul nelocal; dar plata merită să fie efectuată numai atunci cînd sîntem pe deplin încredințați de inevitabilitatea necazului principal: belșugul de extreme locale.**

În sfîrșit, să mai amintim de o metodă de a „cuprinde necuprinsul", care este o perfecționare a metodei celei mai simple, alegerea la întîmplare fără preferințe.

Să împărțim domeniul căutării în două părți egale, notate cu 0 și 1 (fig. 11.2 *a*), să alegem cu totul la întîmplare un punct din domeniul 0 și să aflăm valoarea corespunzătoare f_0. Apoi tot la întîmplare să alegem un punct din domeniul 1; fie f_1 valoarea funcției în acest punct. Să com-

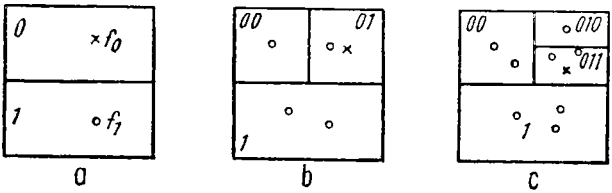


Fig. 11.2

parăm acum f_0 și f_1. Dacă $f_0 > f_1$, împărțim domeniul 0 în două părți cu arie egală, pe care le vom denumi 00 și 01. Dacă $f_1 > f_0$, vom împărți domeniul 1 în două*, obținînd domeniile 10 și 11. Dacă $f_1 = f_0$, lăsăm totul ca mai înainte. Pentru fixarea ideilor să presupunem că $f_0 > f_1$ și că descompunerea s-a efectuat astfel cum se arată în figura 11.2 *b*.

Să facem acum următoarea serie de măsurători: în fiecare dintre domeniile demarcate alegem cu totul la intîmplare cîte un punct pentru evaluarea valorilor. Domeniul în care se constată punctul mai bun (ținînd seama de măsurătorile anterioare), îl vom descompune iarăși pe din două, după care în fiecare dintre domenii vom alege din nou cîte un punct și așa mai departe.

Fondul procedurii este clar: de fiecare dată numărul punctelor măsurătorii crește cu unul și această „creștere" revine domeniului în care anterior a fost constatat un punct mai bun.

Prin aceasta ne asigurăm că

1) nu există vreun punct în domeniul căutării care într-un rînd de măsurători dat să nu poată fi punct de măsurare, și

2) de fiecare dată crește densitatea măsurătorilor în domeniul care a fost recunoscut ca mai bun în virtutea măsurătorilor anterioare.

---

* Împărțirea unui domeniu în două domenii de aceeași mărime este o operație care nu-i nici univocă și nici simplă.

Prima condiție este universală. Ea garantează caracterul global al căutării: în principiu este imposibil să se omită vreun punct, prin urmare nu poate fi omis nici punctul de extremum. Această condiție este satisfăcută de toate metodele descrise în paragraful de față și fireşte că nu este satisfăcută de schemele de căutare locală. A doua condiție caracterizează metoda de „învățare" adoptată în schema respectivă, felul în care se ia în considerare preistoria atunci cînd se alege un rînd de măsurători. Corespunzător cu aceasta, schema descrisă și altele asemănătoare poartă de obicei denumirea *de metoda alegerii la întîmplare cu dirijarea densității măsurătorilor* *.

Așadar am făcut cunoștință cu cinci „metode de a cuprinde necuprinsul".

Fiecare dintre ele (cu eventuala excepție a primei, cu totul „obtuză") își are meritele ei. Toate sînt tinere, suficient de suple și, ceea ce este deosebit de important, insuficient „pătite" în competițiile de căutare.

Lucrul nu e de mirare, căci abia relativ recent a devenit actuală însăși problema căutării locale, astfel încît experiența în materie de rezolvări practice este cu totul redusă.

În treacăt fie zis, ar fi necinstit să ascundem cititorului că unii oameni de știință au în această privință cu totul alt punct de vedere; ei consideră căutarea globală ca fiind tot atît de veche ca și viața însăși. „Într-adevăr — spun ei —, în ce constă schema fundamentală a evoluției materiei vii? Fiecare specie reprezintă un experiment organizat de marele căutător Natura. Mutațiile care apar sînt pașii la întîmplare săvîrșiți în căutarea unor organisme mai bune, mai adaptate. Stabilizarea sau pieirea mutanților în procesul luptei pentru existență este compararea rezul-

---

* Mai răspîndită este denumirea *estimarea aposteriori a densității probabilităților*. Într-adevăr, la început nu există nici un temei pentru a prefera o porțiune a domeniului căutării, astfel încît este firesc ca distribuția inițială, apriorică a probabilităților de alegere să fie uniformă (alegerea este cu totul la întîmplare), dar ulterior, aposteriori, compararea rezultatelor măsurătorilor induce o preferință, astfel încît se poate construi o anumită distribuție a probabilităților diferită de distribuția uniformă. Există mai multe metode raționale pentru asemenea construcții.

tatelor și alegerea celui mai bun dintre ele. Foarte rarele schimbări revoluționare în cadrul cărora apar specii cu totul noi superioare celor anterioare sînt tocmai rarele consecințe fericite ale unor pași mari la întîmplare".

Deși foarte ispititoare, asemenea analogii sînt totuși superficiale. Natura nu este deloc potrivită în rolul de organizator al unei căutări, deoarece ea nu are un țel clar, bine precizat; pe de altă parte, speciile există nu separat, ci într-o complicată interdependență.

În natură se mențin simultan o mulțime de rezultate ale experimentelor, fiecare din ele fiind în felul său reușit și nereușit. Evoluția se săvîrșește în cadrul unei mari varietăți de condiții exterioare care există pe planetă în fiecare moment dat și depinde de varietatea și mai mare a condițiilor existente în trecut.

Astfel încît compararea evoluției cu căutarea locală pare mai curînd doar o metaforă reușită: căutarea globală este în întregime o creație a rațiunii umane și încă o creație foarte tînără *.

---

* Ținînd seama că noțiunile de teoria probabilităților au fost date aici într-un volum extrem de restrîns, autorul nu s-a putut decide să dea descrierea formală a unor proceduri de căutare. Cititorul care se interesează de aceste lucruri și are o pregătire suficientă în domeniul teoriei probabilităților poate să găsească descrierile formale ale procedurilor amintite în acest capitol, de exemplu, în cărțile [1], [22].

12
Cea mai bună căutare a ceea ce este mai bun

(Căutarea unidimensională optimă)

... Vălurile au fost smulse.
Sub lumina crudă, cît e de
simplu totul!
Nimic în lumea aceasta nu
este mai înspăimîntător decît
simplitatea.
E. Vinokurov

După o ascensiune în munți, oamenii de la șes se întorc de obicei entuziasmați: „Vai, ce piscuri înalte, ce văi adînci! Pe culme te orbește soarele strălucitor, în fundul văii te simți parcă înghițit de întuneric. Ce impresii zguduitoare!"

Dar și cel mai pătimaș amator de munți cunoaște acea autentică bucurie care îl cuprinde cînd, după ce a cutreierat haosul muntos, revine la peisajul obișnuit, atît de simplu și familiar.

După ce am rătăcit prin tot felul de multidimensiuni și multiextremități, să revenim la problema aparent cea mai simplă și de aceea cea mai bogată, căutarea unidimensională, pentru a găsi în cercetarea ei atît vechea claritate cît și idei noi.

Iar din problemele unidimensionale vom alege pe cea mai simplă și mai bine organizată. Să presupunem că domeniul căutării este mărginit de un segment de lungime L_0 de la x_{inf} la x_{sup} *. Se știe din capul locului că în acest

* Firește, originea axelor de coordonate poate fi aleasă astfel încît $x_{inf} = 0$ și $x_{sup} = L_0$.

domeniu avem un singur maxim local, deci el este și maxim global.

Mai mult, graficul funcției (fig. 12. 1) nu prezintă porțiuni orizontale, deși sînt posibile frînturi și chiar salturi bruște. Desigur, coordonata maximului, x_m, nu este cunoscută, dar în schimb în orice punct x se poate măsura absolut exact și fără nici un fel de erori valoarea $f(x)$.

Evident, aceasta este o idealizare care simplifică situațiile reale; în schimb, totul este de o extremă claritate.

Cititorul s-ar putea întreba dacă în genere merită să ne ocupăm cu o asemenea problemă, avînd în vedere că avem la dispoziție atîtea metode de rezolvare: și căutarea pe rețea, și căutarea proporțională, și metoda lui Newton, care a perfecționat-o... În cursul a două paragrafe, le-am tot analizat, le-am studiat meritele și lipsurile relative.

Și totuși merită s-o facem, deoarece, trecînd de la o metodă la alta mai bună, am rămas totuși datori cu răspunsul la întrebarea: care metodă este cea mai bună și există oare în genere o metodă optimă de căutare a extremumului? Iată o problemă deosebit de complicată chiar în cazul cel mai simplu.

În primul rînd trebuie să ne înțelegem care metodă putem s-o considerăm optimă, întrucît **orice problemă de optimizare trebuie să înceapă obligatoriu cu formularea funcției-scop, adică cu stabilirea unui criteriu de calitate.**

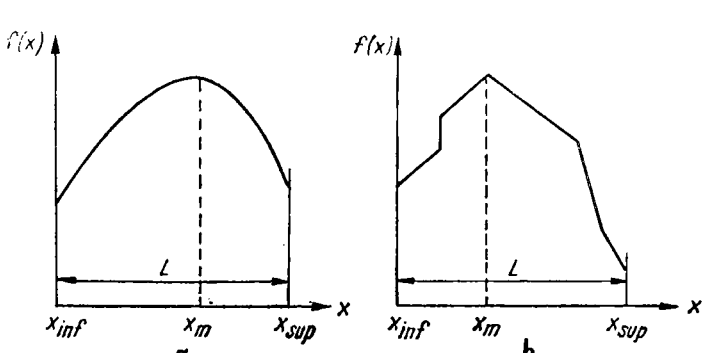


Fig. 12.1

De fapt, deși acest criteriu nu figura în mod explicit, noi îl aveam mereu în vedere: consideram drept mai bună metoda care pretinde un număr mai mic de măsurători pentru găsirea extremumului.

Vorbind mai exact, noi nu am resimțit necesitatea de a descoperi chiar extremumul: ne era suficient să găsim un punct a cărui distanță la extremum să nu depășească un număr Δ suficient de mic.

Acțiunea diferitelor metode de căutare poate așadar să fie privită astfel: la început, înainte de măsurători, poziția extremumului este nedeterminată, el se poate afla în orice punct al domeniului căutării, dar în cursul procesului noi verificăm diferitele puncte și, eliminîndu-le, restrîngem domeniul în care se poate afla realmente extremumul căutat.

De aceea se poate considera drept optimă o metodă care sau

A) permite restrîngerea domeniului inițial de nedeterminare de lungime l_0 pînă la domeniul de lungime dată Δ în urma unui număr minim de măsurători, sau

B) permite restrîngerea domeniului cu lungime dată l_0 în cazul unui număr dat de măsurători N pînă la dimensiunile cele mai mici posibil l_N.

Ambele formulări sînt rezonabile, dar prima (A) este mai firească: aici „omul care caută" poate să-și permită orice cheltuieli pentru a obține precizia dorită, dar el urmărește economisirea la maximum a mijloacelor; formularea (B) se potrivește cazului cînd „omul care caută" dispune de mijloace limitate, dar vrea să-și utilizeze posibilitățile la modul optim.

Acum să încercăm să descriem rezolvarea riguroasă a problemei (B) și să dăm metode aproximative, dar practic eficiente pentru rezolvarea problemei (A); numai că, înainte de a trece la aceasta, să mai precizăm un lucru. Este bine cunoscut că succesul căutării depinde în mod esențial de aspectul concret al funcției-scop. Clasa funcțiilor la care ne limităm (asemenea funcții se numesc de obicei *unimodale*) este totuși suficient de mare: cerințele formulate sînt satisfăcute de curbe foarte diferite (fig. 12.1 *a*, *b*).

Pentru a da o formulare cît mai clară a problemei (B), primul care a studiat-o, savantul american Kiefer, a intro-

dus în formularea ei o cerință suplimentară: dimensiunea domeniului de nedeterminare l_N trebuie să fie garantată pentru orice funcție unimodală, indiferent de care anume funcție particulară ne-ar sili să ne ocupăm vrăjmașa Natură. El a propus să se modifice însuși caracterul problemei (B), dîndu-i-se un alt sens, care, de fapt, este echivalent:

C) Să presupunem că se pot face numai N măsurători și se cere ca în urma acestor măsurători domeniul de nedeterminare în care se poate afla maximul să aibă o lungime nu mai mare decît unitatea. Este necesar să se indice un asemenea număr F_N încît aceasta să fie realizabil pentru orice funcție unimodală dată pe un segment de lungime mai mică decît F_N și să se dea regula de distribuire a punctelor de măsurat care să permită întotdeauna obținerea țelului.

Este limpede că dacă problema (C) este rezolvată, atunci este rezolvată și problema (B): pentru aceasta este suficient să se modifice scara, luînd unitatea de măsură de α ori mai mare decît anterior $\left(\alpha = \frac{l_0}{F_N}\right)$, în care caz în unitățile noi segmentul inițial va avea lungimea $F_N = \frac{l_0}{\alpha}$.

Dar, deși este echivalentă cu problema (B), problema (C) este avantajoasă din punct de vedere matematic. Se constată că numerele F_N se supun unei legități foarte simple

$$F_N = F_{N-1} + F_{N-2},$$

adică fiecare din ele este egal cu suma celor două precedente, iar $F_0 = F_1 = 1$, astfel încît oricare ar fi N aceste numere se calculează lesne:

$$F_2 = F_0 + F_1 = 1 + 1 = 2,$$
$$F_3 = F_1 + F_2 = 1 + 2 = 3$$

și așa mai departe *.

* Acest șir remarcabil de numere întregi a fost construit încă în secolul al XIII-lea de către matematicianul italian Fibonacci în legătură cu o problemă cu totul diferită. Despre istoricul acestei descoperiri, precum și despre numeroasele proprietăți utile ale numerelor F_N, *numerele lui Fibonacci*, se poate citi în broșura lui N.N. Vorobiev [26].

La început numerele F_N cresc încet, dar cu cît N devine mai mare, cu atît creșterea lor devine mai impetuoasă și F_{20} depășește 10 000. Adică cu ajutorul doar a 20 de măsurători segmentul inițial de nedeterminare se poate reduce la o sutime din lungimea sa inițială!

E drept că încă nu știm cum să planificăm distribuirea punctelor de măsurat; și nici n-am demonstrat încă faptul că lungimea segmentului F_N este supusă legității pe care am enunțat-o.

Să încercăm mai întîi să dăm această demonstrație. Vom face apel la metoda inducției matematice, dar într-o variantă destul de ingenioasă.

În primul rînd, este clar că dacă nu efectuăm nici o măsurătoare ($N = 0$) sau dacă efectuăm numai una singură ($N = 1$), nedeterminarea n-are cum să dispară; dacă segmentul inițial a avut o lungime mai mare decît unitatea, atunci el așa va rămîne. Prin urmare, $F_0 = F_1 = 1$.

Să studiem acum cazul în care dispunem numai de două măsurători ($N = 2$). După cum vom vedea ulterior, acest caz merită o atenție cu totul deosebită.

Așadar, să presupunem că segmentul inițial are o lungime L (fig. 12.2). Să efectuăm măsurători în două

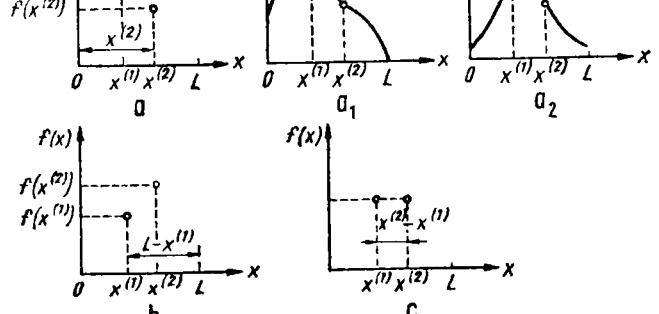


Fig. 12.2

puncte deocamdată arbitrare $x^{(1)}$ și $x^{(2)}$, unde pentru fixarea ideilor (ordinea numerotării fiind arbitrară) presupunem $x^{(2)} > x^{(1)}$. Ca întotdeauna sînt posibile trei rezultate (fig. 12.2 *a*, *b*, *c*)

$$\text{a) } f(x^{(1)}) > f(x^{(2)}), \text{ b) } f(x^{(1)}) < f(x^{(2)}), \text{ c) } f(x^{(1)}) = f(x^{(2)}).$$

Să analizăm fiecare din aceste trei cazuri, fără a uita că funcția de studiat este obligatoriu unimodală.

Cazul (a) este posibil fie (a_1) dacă ambele puncte se află pe versantul drept al graficului funcției f, adică la dreapta extremumului, fie (a_2) dacă punctele sînt situate pe versanți diferiți, de ambele părți ale extremumului.

A treia situație ipotetică (ambele puncte pe versantul stîng) este evident imposibilă în cazul unui asemenea rezultat al măsurătorilor. Totodată este clar că în subcazul (a_1) extremumul se poate afla numai pe segmentul dintre 0 și $x^{(1)}$, iar în subcazul (a_2) el trebuie să se afle pe segmentul dintre $x^{(1)}$ și $x^{(2)}$. Ce s-a întîmplat în realitate nu știm, dar cel puțin este clar că dacă în urma măsurătorilor avem

$$f(x^{(1)}) > f(x^{(2)}),$$

atunci coordonata maximului nu poate fi la dreapta lui $x^{(2)}$; așadar, în cazul unui asemenea rezultat al măsurătorilor dimensiunea domeniului nedeterminat a devenit egală cu $x^{(2)}$.

Un raționament analog arată că în cazurile de tip (b) extremumul nu poate fi la stînga lui $x^{(1)}$. Prin urmare, lungimea nedeterminării rămîne egală cu $L - x^{(1)}$. În sfîrșit, un rezultat ca cel de la cazul (c) poate avea loc numai dacă punctele $x^{(1)}$ și $x^{(2)}$ sînt situate de ambele părți ale maximului (întrucît pe versanți nu putem avea porțiuni orizontale). Prin urmare, în cazul (c) zona de nedeterminare este de lungime $x^{(2)} - x^{(1)}$.

Așadar, putem alcătui un tabel în care pentru fiecare caz al rezultatelor măsurătorilor să indicăm dimensiunea noului domeniu de nedeterminare.

a	b	c
$x^{(2)}$	$L - x^{(1)}$	$x^{(2)} - x^{(1)}$

Dar, indiferent ce anume variantă are loc în realitate și independent de aspectul concret al funcției studiate, noi trebuie să garantăm ca această dimensiune să nu fie mai mare decît unitatea. Prin urmare, punctele $x^{(1)}$ și $x^{(2)}$ trebuie alese astfel încît să fie satisfăcute concomitent toate cele trei condiții:

$$x^{(2)} \leqslant 1, \quad L - x^{(1)} \leqslant 1, \quad x^{(2)} - x^{(1)} \leqslant 1.$$

E drept, a treia condiție este întotdeauna îndeplinită dacă este îndeplinită prima; așadar ea poate fi înlăturată și ne rămîn condițiile

$$1)\ x^{(2)} \leqslant 1, \qquad 2)\ L \leqslant 1 + x^{(1)},$$

care, desigur, trebuie completate cu restricțiile naturale

$$0 \leqslant x^{(1)} \leqslant L \quad (1); \quad 0 \leqslant x^{(2)} \leqslant L \quad (2); \quad x^{(2)} > x^{(1)} \quad (3).$$

Să examinăm planul axelor de coordonate, așa cum este indicat în figura 12.3. Pe axa absciselor vom fixa coordonata primului punct măsurat $x^{(1)}$, iar pe axa ordonatelor — coordonata celui de-al doilea punct, $x^{(2)}$. Fiecare punct din plan corespunde unei anumite perechi de asemenea coordonate, dar de la început este limpede că în virtutea restricțiilor (1) și (3) ne interesează numai punctele situate în triunghiul haşurat, și încă punctele de pe ipotenuza sa sînt inadmisibile (dacă $x^{(1)} = x^{(2)}$, atunci ambele puncte se confundă și nu putem obține nici o informație); prin urmare, $x^{(1)}$ trebuie să fie mai mic decît 1.

Pe de altă parte, în virtutea restricției (2), lungimea segmentului inițial nu depășește $1 + x^{(1)}$.

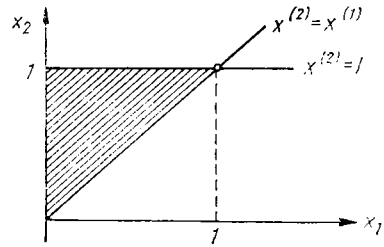


Fig. 12.3

Prin urmare, L nu este mai mare decît 2 și poate fi făcut infinit de apropiat de această mărime dacă luăm $x^{(2)} = 1$, iar $x^{(1)}$ oricît de apropiat de numărul 1, dar fără a coincide exact cu $x^{(2)}$ *.

Așadar 2 este marginea superioară a lungimii posibile a segmentului inițial de nedeterminare, care poate fi redus la segmentul unitate în N măsurători, dacă $N = 2$. Prin aceasta am demonstrat că

$$F_2 = 2;$$

dar $F_0 = F_1 = 1$; rezultă:

$$F_2 = F_1 + F_0.$$

Iată deci că ipoteza

$$F_N = F_{N-1} + F_{N-2}$$

s-a dovedit adevărată pentru $N = 2$. Rămîne să folosim inducția după N pentru a demonstra că ea este adevărată întotdeauna.

Urmînd standardele clasice, să presupunem că pentru $k = 2,3, \ldots, N-1$ avem

$$F_k = F_{k-1} + F_{k-2}$$

și să demonstrăm valabilitatea acestei afirmații pentru $k = N$.

---

* Ipoteza inițială cu privire la posibilitatea unor măsurători absolut exacte admite compararea rezultatelor în puncte oricît de învecinate.

Să examinăm din nou segmentul de lungime L și să ne gîndim cum să dispunem primele două măsurători din cele N măsurători pe care avem să le efectuăm în vederea reducerii lui L la un segment de lungime nu mai mare decît unitatea. Să convenim că L nu este atît de mare încît acest lucru să fie imposibil de făcut în cazul unei planificări raționale a măsurătorilor. Raționalitatea, firește, trebuie să aibă loc de la bun început, de la primele două măsurători; ele trebuie organizate astfel încît să nu strice posibilitățile ulterioare.

Coordonatele acestor prime două măsurători le vom nota cu $x^{(1)}$ și $x^{(2)}$; să considerăm iarăși toate cazurile rezultatelor posibile. Este posibil cazul $f(x^{(1)}) > f(x^{(2)})$ indicat în figura 12.2 *a*. Evident că în acest caz nu avem dreptul de a alege $x^{(2)} > F_{N-1}$. Într-adevăr, domeniul poziției posibile a maximului s-a redus la un segment de lungime $x^{(2)}$, în interiorul căruia în punctul $x^{(1)}$ s-a efectuat deja o măsurătoare. Pe acest segment putem distribui cele $N-2$ măsurători care ne mai stau la dispoziție, dar cu $N-1$ măsurători nu se poate reduce la lungimea unitate un segment mai mare decît F_{N-1}. Prin urmare, pentru a nu strica definitiv posibilitățile de viitor, este necesar ca a doua măsurătoare s-o alegem astfel încît să fie satisfăcută condiția $x^{(2)} \leqslant F_{N-1}$. În plus, trebuie să ținem seama că nu este exclusă situația prezentată de figura 12.2*a*, cînd maximul se află pe segmentul dintre 0 și $x^{(1)}$, în interiorul căruia nu s-a efectuat nici o măsurătoare. Deci acest segment va trebui redus la unul de lungime unitate cu ajutorul celor $N-2$ măsurători care ne-au rămas la dispoziție. Prin urmare, este necesar să asigurăm ca $x^{(1)} \leqslant F_{N-2}$.

Pe de altă parte, este posibil cazul (b), cînd $f(x^{(1)}) < f(x^{(2)})$ și realizarea maximului e posibilă undeva pe segmentul de lungime $L - x^{(1)}$ (fig. 12.2 *b*). Printr-un raționament asemănător cu cel anterior este lesne de stabilit că în această situație este necesar să supunem $x^{(1)}$ restricției exprimată de condiția $L - x^{(1)} \leqslant F_{N-1}$.

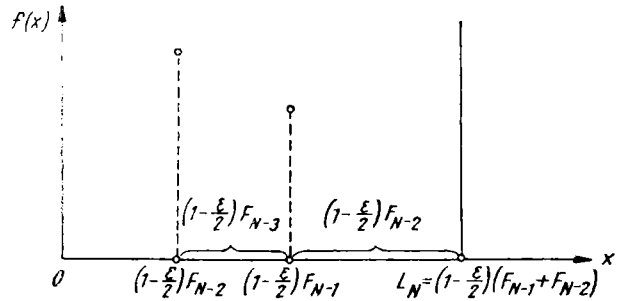


Fig. 12.4

Or, toate aceste situații enumerate * sînt posibile; prin urmare, pentru a fi asigurați în viitor, trebuie să impunem la alegerea coordonatelor primelor două puncte îndeplinirea simultană a tuturor condițiilor de mai sus:

$$x^{(2)} \leqslant F_{N-1},\ x^{(1)} \leqslant F_{N-2},\ L - x^{(1)} \leqslant F_{N-1}.$$

Din aceste condiții rezultă că ele pot avea loc numai dacă

$$L \leqslant F_{N-1} + x^{(1)} \leqslant F_{N-1} + F_{N-2}.$$

Prin urmare, cea mai mare lungime posibilă a segmentului inițial de nedeterminare care poate fi redusă pînă la un segment de lungimea unitate uzînd de N măsurători nu poate fi mai mare de $F_{N-1} + F_{N-2}$, cu alte cuvinte

$$F_N \leqslant F_{N-1} + F_{N-2}.$$

Iar coordonatele măsurătorilor inițiale trebuie, în particular, să satisfacă condițiile

$$x^{(1)} \leqslant F_{N-2},\ x^{(2)} \leqslant F_{N-1}.$$

Dispunînd de aceste informații utile, să examinăm acum (fig. 12.4) segmentul de lungime

$$L = L_N = \left(1 - \frac{\varepsilon}{2}\right)(F_{N-1} + F_{N-2}),$$

* Examinarea cazului (c) nu aduce nimic nou.

unde ε este un număr pozitiv arbitrar de mic și să situăm punctele primelor două măsurători în felul următor:

$$x^{(1)} = \left(1 - \frac{\varepsilon}{2}\right) F_{N-2}, \quad x^{(2)} = \left(1 - \frac{\varepsilon}{2}\right) F_{N-1};$$

adică le așezăm simetric la distanțe egale de extremități. În felul acesta am satisfăcut toate condițiile care garantează succesul final și totodată am dovedit că acest lucru se poate face oricît de apropiată de $F_{N-1} + F_{N-2}$ ar fi lungimea L a segmentului inițial. Dar, prin definiție, F_N este de așa natură încît orice segment de lungime mai mică decît F_N desigur că poate fi redus la un segment de lungimea unitate. Așadar, în definitiv am obținut că

$$F_N = F_{N-1} + F_{N-2}.$$

În felul acesta demonstrația inductivă a fost încheiată, întrucît valabilitatea ei pentru $N = 2$ a fost stabilită anterior. Lucrul cel mai important este că în cursul demonstrației am stabilit și regula care ne dă cea mai bună metodă a căutării. S-o descriem mai amănunțit.

Luînd segmentul inițial de lungime $L_N = \left(1 - \frac{\varepsilon}{2}\right) F_N$, unde ε poate fi oricît de mic dorim (în practică micimea lui ε este limitată numai de discernabilitatea rezultatelor măsurătorilor), să așezăm primele două puncte simetric la distanța de extremități $\left(1 - \frac{\varepsilon}{2}\right) F_{N-2}$. Din condiția $F_N = F_{N-1} + F_{N-2}$ și din figura 12.5 *a,b* rezultă că, oricare ar fi rezultatul măsurătorii, segmentul de nedeterminare se va reduce la un segment de lungime $L_{N-1} = \left(1 - \frac{\varepsilon}{2}\right) F_{N-1}$, în interiorul căruia, la distanța $\left(1 - \frac{\varepsilon}{2}\right) F_{N-3}$ de extremitatea dreaptă sau stîngă, se află punctul în care a fost efectuată măsurătoarea și s-a obținut rezultatul mai bun (din cele două aflate). Executînd o a treia măsurătoare simetrică cu aceasta (vezi punctul marcat în fig. 12.5 *a, b*),

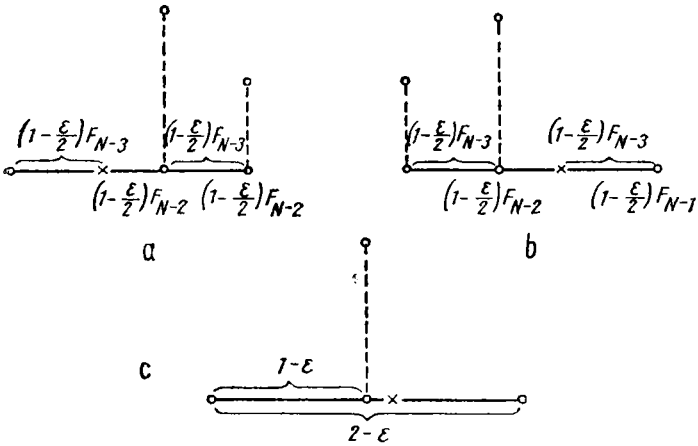


Fig. 12.5

vom obține o reducere pînă la $L_{N-2} = \left(1 - \frac{\varepsilon}{2}\right) F_{N-2}$. Menținînd aceeași ordine de acțiune, după a $(N\text{-}1)$-a măsurătoare vom ajunge la situația cînd avem un segment de lungime $L_2 = \left(1 - \frac{\varepsilon}{2}\right) F_2 = \left(1 - \frac{\varepsilon}{2}\right) \cdot 2 = 2 - \varepsilon$ cu o măsurătoare executată deja în interiorul său și oferind cea mai bună din toate valorile funcției găsite anterior (fig. 12.5*c*). Ultima măsurătoare, a N-a, fixează segmentul de lungime unitate în care poate fi situată coordonata maximului și dă o estimație inferioară a valorii maxime a funcției-scop, întrucît s-a păstrat cea mai bună dintre valorile măsurate anterior.

Amintim încă o dată că dacă mărimea domeniului de căutare este egală cu un număr l_0, atunci printr-o modificare corespunzătoare a scării, adică prin introducerea unei unități noi mai mari de $\alpha = \frac{l_0}{F_N}$ ori se poate realiza reducerea la algoritmul trebuincios. Mai exact, ar trebui luat $\alpha = \frac{l_0}{F_N(1-\varepsilon/2)}$, dar avînd în vedere că ε

este presupus mic, rezultatul aproape că nu va fi modificat. Segmentul de nedeterminare la care ajungem după N măsurători va avea în unități inițiale lungimea $l_N = \alpha$. Desigur, pentru a ajunge la scara dorită trebuie să cunoaștem numerele F_N. Să le calculăm în prealabil pînă la $N = 20$ și să le trecem într-un tabel. În același tabel vom trece raporturile a două numere consecutive F_N/F_{N-1} care caracterizează cîștigul în reducerea segmentului de nedeterminare atunci cînd se admite o măsurătoare în plus. Constatăm că acest raport începînd cu $N = 6$ devine, cu două zecimale exacte, 1,62 și ulterior, cînd N crește, raportul rămîne constant. Teoretic, limita către care tinde acest raport va fi

$$\frac{1 + \sqrt{5}}{2} = 1{,}61803\ldots$$

N	0	1	2	3	4	5	6	7	8	9	10
F_N	1	1	2	3	5	8	13	21	34	55	89
F_N/F_{N-1}		1,00	2,00	1,50	1,66	1,60	1,62	1,62	1,62	1,62	1,62
N		11	12	13	14	15	16	17	18	19	20
F_N		144	233	377	610	987	1597	2584	4181	6765	10946
F_N/F_{N-1}		1,62	1,62	1,62	1,62	1,62	1,62	1,62	1,62	1,62	1,62

Acesta este un fapt foarte important și util: într-adevăr procedura optimă de căutare, așa cum am descris-o, are o deficiență practic esențială, deoarece pentru a o realiza trebuie să știm dinainte de care anume număr de măsurători dispunem. Ce ne facem dacă în cursul căutării se lămurește că mai sînt necesare doar cîteva măsurători? Atunci, în genere vorbind, va rezulta că întregul proces anterior ar trebui revizuit, deoarece procedura optimă pentru $N + 2$ măsurători diferă datorită modului în care s-au organizat primele N măsurători de procedura optimă pentru N dat în prealabil.

Dar să revenim la figurile 12.4 și 12.5. Examinîndu-le, vedem lămurit că de fapt regula de distribuție depinde numai de raportul $\frac{F_k}{F_{k-1}}$ ($k = N, N-1, \ldots$) dintre lungimea totală a segmentului în etapa respectivă și distanța la extremitatea dreaptă, respectiv stîngă a celor două puncte simetrice așezate în interiorul său. Or, am văzut că pentru k suficient de mare ($k \geqslant 6$) acest raport este constant și aproximativ egal cu 1,62. Prin urmare, dacă s-a efectuat un număr N suficient de mare de măsurători și este de dorit să se efectueze încă un număr de măsurători, nu mai mic de șase (iar în practică nu mai mic de trei), atunci întreaga preistorie a căutării poate fi considerată drept optimă și pentru această mulțime de măsurători, care este mai mare.

Mai mult, apare încă o idee simplă: întrucît căutarea optimă care se întemeiază pe numerele lui Fibonacci dă la fiecare pas suficient de îndepărtat de cel anterior o reducere a segmentului de nedeterminare de unul și același număr de ori, nu se poate oare simplifica procedura, respectînd același principiu, dar la toți pașii?

Să arătăm că acest lucru este posibil și deloc complicat. Să presupunem că pe segmentul de lungime L în care se poate situa coordonata maximului s-au efectuat două măsurători în punctele $x^{(1)}$ și $x^{(2)}$. Atunci reducerea garantată a segmentului va fi egală cu cel mai mic dintre numerele

$$\frac{L}{x^{(2)}} \quad \text{și} \quad \frac{L}{L - x^{(1)}}.$$

Să așezăm punctele simetric; atunci ambele numere vor fi egale cu o anumită mărime τ (fig. 12.6 *a*).

$$\frac{L}{x^{(2)}} = \tau \text{ și } \frac{L}{L - x^{(1)}} = \tau$$

sau

$$x^{(2)} = \frac{L}{\tau}, \; L - x^{(1)} = \frac{L}{\tau}.$$

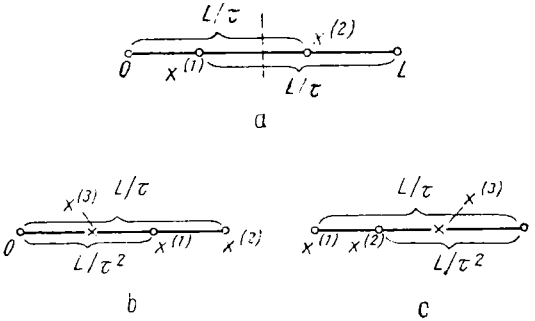


Fig. 12.6

În felul acesta se garantează că, în urma măsurătorilor, lungimea segmentului de nedeterminare se va reduce de τ ori, adică pînă la $\frac{L}{\tau}$, iar în interiorul său va rămîne o măsurătoare situată la distanța

$$x^{(1)} = L - x^{(2)} = L - \frac{L}{\tau}$$

de una din extremități. Să convenim a alege τ astfel încît această distanță să fie de τ ori mai mică decît lungimea totală rămasă, adică cerem ca τ să satisfacă condiția

$$L - \frac{L}{\tau} = \frac{L}{\tau^2} \text{ sau } \tau^2 - \tau - 1 = 0.$$

Atunci noua măsurătoare $x^{(3)}$ se va așeza simetric față de cea disponibilă, adică la aceeași distanță $\frac{L}{\tau^2}$ socotită de la cealaltă extremitate (12.6 *b*, *c*), ceea ce garantează iarăși reducerea nedeterminării de τ ori. Este evident că o asemenea procedură se poate repeta în oricît de mulți pași, garantînd de fiecare dată reducerea exact de τ ori. Rămîne numai să calculăm care anume τ satisface condi-

ția pusă. Or, această condiție este o ecuație de gradul al doilea în τ și ca atare ne oferă două rădăcini

$$\frac{1+\sqrt{5}}{2} \quad \text{și} \quad \frac{1-\sqrt{5}}{2}.$$

Pe noi desigur că ne interesează prima rădăcină, cea pozitivă; dar aceasta este tocmai numărul cunoscut 1,62, care caracteriza posibilitățile procedurii Fibonacci optime pentru N mare. Așadar, o regulă de o simplitate elementară — regula conform căreia de fiecare dată măsurătorile se efectuează în puncte simetrice așezate la distanța $1/1{,}62 \approx$ $\approx 0{,}62$ de extremitățile segmentului de nedeterminare rămas — se comportă, în cazul unui număr mare de măsurători care se mai pot încă efectua nu mai rău decît savanta procedură optimă. Chiar dacă o menținem în vigoare pînă la sfîrșit, cînd numărul măsurătorilor ce urmează a mai fi efectuat devine mic, paguba (în comparație cu căutarea optimă) este neînsemnată.

Cu N măsurători se poate obține o reducere a segmentului inițial de $(1{,}62)^{N-1}$ ori (prima măsurătoare nu generează nici o reducere). Se poate dovedi că acest număr nu diferă niciodată de F_N cu mai mult de 23% (cazul cel mai prost este $N = 3$), iar pentru un N mare diferența va fi de numai 17% *.

Fiind independentă de N, firește că această regulă este adecvată și ca rezolvare a problemei (A), în cadrul căreia numărul de măsurători nu este fixat dinainte; iar faptul că ea diferă puțin ca eficiență de regula optimă pentru N cunoscut ne permite s-o considerăm aproximativ optimă și în cazul problemei (A).

Regula are o denumire răsunătoare: *metoda secțiunii de aur*. Originea acestei denumiri trebuie căutată în

* În plus, regula este lipsită de deficiența practică a procedurii optime care impune necesitatea organizării ultimelor măsurători la o distanță mică una de alta; aici însă, distanța rămîne întotdeauna egală cu $\frac{2}{\tau} - 1 \cong 0{,}24$ din lungimea totală a segmentului.

timpuri străvechi, la marele geometru grec Euclid. *Într-adevăr, regula descrisă constă în efectuarea unei măsurători exact în punctul care împarte un segment dat astfel încît raportul dintre lungimea L a întregului segment și lungimea subsegmentului mai mare să fie egal cu raportul dintre acesta din urmă și subsegmentul mai mic.*

$$\frac{L}{L-x^{(1)}} = \frac{L-x^{(1)}}{x^{(1)}} = \tau;$$

or, problema de a se împărți astfel un segment (cu ajutorul riglei și compasului!) este amintită în *Elementele* lui Euclid sub denumirea „Problema secțiunii de aur". Fie ca regula „secțiunii de aur" să încununeze eforturile noastre în căutarea procedurilor de căutare optime.

„Dar — va spune cititorul, căruia acum i s-a trezit gustul pentru acest gen de probleme — n-am studiat decît problema unidimensională cea mai simplă; n-ar fi cazul să încercăm să trecem la probleme multidimensionale, transpunînd acolo noile idei cu care am făcut cunoștință?" Din păcate, această cale, străbătută cu atîta succes în cadrul metodelor locale de căutare, aici este aproape total lipsită de orice speranță. Dealtfel, cititorul atent probabil că a și înțeles acest lucru: nenorocirea constă în aceea că orice punct divide un segment rectiliniu în două părți, pe cînd în plan un singur punct, o singură măsurătoare, nu divide nimic. De aceea și clasificarea domeniilor despărțite prin diferite măsurători, clasificare pe care se întemeiază rezolvarea optimă a problemelor unidimensionale, este, în general vorbind, imposibilă chiar în cazul bidimensional *.

Firește, de aici nu trebuie să tragem concluzia pripită că cele discutate mai sus sînt cu totul inutile pentru rezolvarea problemelor multidimensionale. Să ne amintim de cea mai renumită dintre metodele locale de căutare, aceea

* Se cunosc încercări de generalizare directă a ideii de căutare optimă la probleme multidimensionale (a se vedea de exemplu [1], dar toate aceste încercări impun restricții nu prea firești comportării posibile a funcțiilor studiate.

a campionului nostru Panta-cea-mai-rapidă. După cum vă amintiți, acțiunea lui avea două componente. La început el determina direcția cea mai bună, gradientul, iar apoi se ocupa cu căutarea unidimensională după direcția aleasă. Această a doua componentă poate fi adesea și foarte eficient îmbogățită prin introducerea noilor idei.

La drept vorbind, putem presupune că arma cea mai secretă a campionului, armă la care făcea aluzie Profesorul cînd se despărțea de noi în cap. 6, este tocmai căutarea optimă după direcție.

Chiar dacă am efectua-o după regula simplificată a „secțiunii de aur", avem șanse să obținem o ridicare considerabilă a eficacității în comparație cu trierea directă pe rețea sau cu căutarea proporțională. *Vorbim aici de „șanse" și nu de „certitudine" din două motive asupra cărora merită să ne oprim.*

Primul motiv rezidă în aceea că funcția unidimensională de studiat nu va fi în mod obligatoriu unimodală. Dar aceasta nu este o pagubă prea mare, deoarece pentru asemenea funcții și alte metode locale își pierd parțial eficacitatea.

Al doilea motiv este mult mai important și derivă din însuși fondul definiției pe care am utilizat-o pentru conceptul de „optim". Doar noi tindem să găsim o metodă care să garanteze rezultatele cele mai bune pentru orice funcție unimodală posibilă, oricît de nepotrivită și incomodă ar fi ea. Dacă însă, în realitate, funcția are un aspect mai cuviincios, atunci alte metode locale, în particular dintre cele pe care le cunoaștem, pot să ne conducă mai rapid la succes.

Să amintim, de exemplu, remarcabila metodă a lui Newton întemeiată pe estimarea înclinării funcției cît și a înclinării înclinării înseși (cap. 5). Cu ajutorul unor grafice ne-am convins că dacă în zona căutării înclinarea înclinării este aproape constantă, atunci metoda lui Newton ne conduce într-o mică vecinătate a extremumului cu ajutorul unui singur pas, care pretindea numai trei măsurători.

Or, cazul cînd înclinarea inclinării (derivata a doua) este constantă în vecinătatea extremumului este un caz destul de frecvent în practica căutării. În asemenea cazuri metoda lui Newton este mai eficientă decît o metodă mai generală, bună pentru tot felul de cazuri imprevizibile.

Desigur, **toate aceste considerente nu pot umbri meritele metodelor optime sau aproape optime, oricare ar fi neplăcerile.** În ultimă instanță este rațional să-ți asiguri casa pentru caz de incendiu, chiar dacă pînă în prezent n-ai avut niciodată parte de un incendiu.

13
Drumurile pe care le alegem

(Programarea dinamică)

Drumul tău
Tu singur ți-l alegi.
Pe margini de prăpastii de se-abate
Sau duce drept la țintă
Tot al tău e,
E drumul ce-ai ales.

E. Akiko

Probabil că încă din copilărie cititorul cunoaște următoarea problemă distractivă: cineva i-a strecurat unui zaraf o monedă falsă și el trebuie s-o descopere într-un maldăr de monede de aur adevărat cu care cea falsă seamănă întru totul. Desigur, asemănarea nu-l poate înșela pe încercatul zaraf: el știe foarte bine că moneda falsă are altă greutate decît cea de aur adevărat. Așadar este suficient să cîntărească fiecare monedă în parte și cea netrebnică va fi descoperită. Da, dar această banală metodă pretinde prea mult timp: zaraful este om de afaceri, adevărul că „timpul este bani" el l-a asimilat prea bine și nu-și poate îngădui să irosească atîta timp de prisos pentru cîntărirea monedelor; va folosi deci o altă metodă, cea mai rapidă dintre toate metodele posibile.

Cum trebuie să procedeze bătrînul zaraf pentru a găsi printre N monede una falsă, aceasta este o problemă cunoscută din copilărie.

Rezolvarea ei nu este dificilă. Ea este însă atît de instructivă încît merită să ne oprim mai mult asupra ei, examinînd-o amănunțit de pe poziții strict științifice.

În primul rînd, să numerotăm în gînd toate monedele de la 1 pînă la N și să ne închipuim că există o dependență între greutatea monedei și numărul ei de ordine (fig. 13.1). Greutatea y depinde de numărul de ordine x, care admite un număr finit de valori. Astfel, dintr-un punct de vedere mai general, căutarea monedei false este căutarea minimului unei funcții care este dată pe o mulțime finită, presupunînd că se cunoaște în prealabil că moneda falsă este mai ușoară. E suficient așadar să facem uz de una dintre numeroasele metode de căutare pe care le cunoaștem pentru a rezolva problema nu mai rău decît zaraful.

Trierea primitivă desigur că o vom respinge, încercînd ceva mai inteligent, de exemplu căutarea locală ; din păcate, compararea cu vecinii nu ne va da nimic, deoarece toți vecinii sînt identici, nimeni nu poate să pretindă că ar fi „mai bun", ceea ce, în cazul de față înseamnă „mai fals" ; prin urmare, vecinii nu ne pot indica direcția căutării. S-ar putea să ne ajute procedura căutării optime, dar și aici eșecul este deplin; funcția reprezentată în figura 13.1 nu e deloc unimodală: atît la dreapta cît și la stînga minimului avem porțiuni orizontale. Compararea măsurătorilor, adică compararea greutăților diferitelor monede, nu va da reducerea dorită a domeniului căutării.

Deruta este astfel totală... Dar e posibil ca știința modernă a căutării să se dea bătută în fața unui zaraf din vremuri imemoriale?

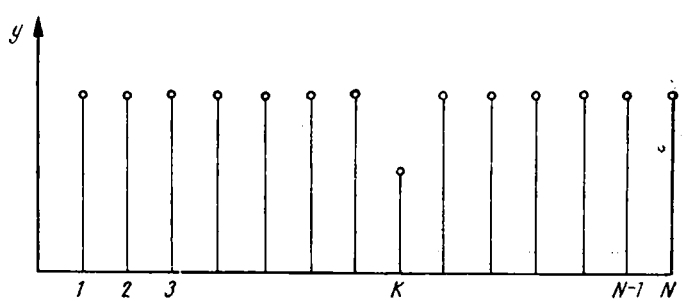


Fig. 13.1

Ei bine, nu; concluzia aceasta ar fi pripită. Am uitat pur și simplu că zaraful putea cîntări nu numai fiecare monedă în parte, dar și cîteva monede dintr-o dată, unindu-le în grupe. Or, în felul acesta lucrurile iau o cu totul altă întorsătură: în noile condiții principiile căutării optime ne vor conduce la succes.

În primul rînd să ne înțelegem asupra unei condiții care nu este de mică importanță: ce anume cîntar are la dispoziție zaraful. Să considerăm pentru început cazul cel mai simplu, cînd zaraful dispune de un cîntar cu arc sau de o balanță romană și să presupunem în plus că el cunoaște în prealabil greutatea p a unei monede veritabile.

Atunci, cîntărind dintr-o dată un număr k de monede, zaraful poate obține unul din următoarele două rezultate:

a) greutatea poate fi egală cu $p \cdot k$, în care caz nici una dintre monedele cîntărite nu este falsă și, prin urmare, moneda falsă se află printre celelalte $N-k$ monede;

b) greutatea monedelor diferă de $p \cdot k$ și, prin urmare, moneda falsă se află tocmai printre cele k monede cîntărite.

Făcînd uz de cunoscutul punct de vedere că totdeauna trebuie să fim gata pentru posibilitatea cea mai rea și, orice situație ar apărea, să ne asigurăm rezultatul cel mai bun, zaraful trebuie să aleagă numărul k al monedelor cîntărite astfel încît dintre cele două numere k și $N-k$ cel mai mare să aibă valoarea minimă. Este clar că dacă N este un număr par, atunci trebuie luat $k = \frac{N}{2}$, adică trebuie cîntărite jumătate dintre monede. Oricare ar fi rezultatul cîntăririi, nedeterminarea monedelor suspecte se reduce în urma cîntăririi la $N/2$ monede. Dacă însă N este impar, atunci luînd $k = \frac{N-1}{2}$, putem ajunge în cazul cel mai rău (cînd toate monedele cîntărite se dovedesc veritabile) la nedeterminarea alegerii printre

$$N - k = N - \frac{N-1}{2} = \frac{N+1}{2}$$

monede. Acesta este algoritmul zarafului! Șirul converge aproape tot atît de repede ca și progresia geometrică. Să presupunem, de exemplu, că avem 23 de monede. Atunci, chiar în cazul rezultatelor celor mai nefavorabile, după fiecare cîntărire numărul monedelor suspecte se va reduce precum urmează. Avem 23. Cîntărim 11. Rezultatul: veritabile. Rămîn 12. Cîntărim 6. Independent de rezultat, rămîn 6. Cîntărim 3. Independent de rezultat, rămîn 3. Cîntărim 1. Rezultat: veritabilă. Rămîn 2. Cîntărim 1 și constatăm care din ultimele două monede este falsă.

Deci: cîte cîntăriri vor fi necesare (în cazul rezultatelor celor mai defavorabile!) pentru a descoperi moneda falsă dintr-un număr arbitrar N de monede? Cititorul se va convinge cu ușurință că dacă N este mai mare decît 2^{m-1} și nu depășește 2^m, vor fi necesare m cîntăriri *. În exemplul considerat am avut $N = 23$, dar $32 = 2^5 > 23 > 2^4 = 16$, astfel încît au fost necesare cinci cîntăriri.

Să examinăm acum cazul mai complicat, cînd greutatea monedei veritabile nu este cunoscută în prealabil, iar zarafful are la dispoziție obișnuita balanță romană. Greutăți nu are de nici un fel, dar nici nu este nevoie de ele: într-adevăr, este necesar doar să descoperim moneda mai ușoară și nu să aflăm cu cîte grame este mai ușoară, deci în loc de greutăți putem recurge la monedele înseși**. Să punem pe un taler al cîntarului k monede și tot atîtea pe celălalt. Dacă se va stabili echilibrul, toate monedele care se află pe talere sînt, evident, veritabile, adică moneda falsă se află printre celelalte $N-2k$ monede. În caz contrar, unul din talere se va ridica, deci moneda falsă se va afla printre cele k monede de pe talerul respectiv. Ca să ne asigurăm pentru condițiile cele mai grele, vom alege pe k

---

* Cu alte cuvinte, numărul cîntăririlor este egal cu numărul cifrelor din transcrierea lui N în sistemul binar. De exemplu, numărului care în sistemul zecimal se notează cu 23 îi corespunde forma binară 10111, care are cinci cifre.

** Este de remarcat că în cazul a două monede, dacă nu se știe care din ele — cea veritabilă sau cea falsă — este mai grea, problema este insolubilă.

astfel încît cel mai mare dintre cele două numere $N-2k$ și k să fie pe cît posibil mai mic.

Să împărțim N la 3: obținem un cît x și, eventual, un rest a, egal cu 1 sau 2:

$$N = 3x + a.$$

Dacă n-avem rest, atunci punem $k = x$ și în urma unei singure cîntăriri am redus numărul monedelor suspecte la x, deoarece $N-2k$ este de asemenea egal cu x. Dacă însă avem rest, atunci luînd $k = x + 1$, garantăm reducerea nedeterminării la $x + 1$, deoarece

$$N - 2k = x + a - 2$$

nu poate fi mai mare decît x. Este limpede că această regulă dă întotdeauna cel mai mare succes garantat și poate fi aplicată iar și iar mulțimii nou formate de elemente suspecte. Ordinea de succesiune a cîntăririlor pentru $N = 23$, în cazul rezultatelor celor mai defavorabile, este dată de următorul tabel:

N	23	8	3
x	7	2	1
k	8	3	1

Pentru descoperirea monedei false au fost necesare doar trei cîntăriri. Propunem cititorului să demonstreze singur că dacă numărul inițial de monede N este mai mare decît 3^{m-1} și nu depășește 3^m, pentru garantarea succesului vor fi necesare m cîntăriri. Cititorul se poate convinge și de faptul că în cazul prezenței restului regula indicată nu este unica:

N	23	9	3
k	9	3	1

În cele din urmă, cititorul se va indigna de-a binelea și va spune: „Pînă acum consideram că în această carte se vorbește neserios despre lucruri serioase, dar acum sînt nevoit să citesc discuții serioase despre o veche ghicitoare neserioasă!"

Dar, stimate cititor de înaltă erudiție, să nu ne grăbim totuși cu concluziile. Să vorbim la început despre o problemă pe deplin serioasă. Într-un circuit electric sînt legate succesiv o mulțime de dispozitive. Deodată unul din ele se deteriorează și circuitul se întrerupe. Aceasta se constată imediat, deoarece lipsește curentul. Va trebui să ne ocupăm cu reparația. Dar la început trebuie să determinăm unde anume a avut loc întreruperea. Firește, am putea verifica totul la rînd, dar avem la dispoziție și o tactică mai rațională. Care anume? Întrebarea este firește întîrziată: doar am analizat amănunțit cea mai bună tactică a zarafului cînd trebuia să caute moneda falsă, iar problema căutării unei dereglări nu diferă în fond prin nimic de problema zarafului, deși cîntărirea este înlocuită cu „încercarea" porțiunilor din lanț (fig. 13.2).

Se pare că cititorul serios a pierdut dreptul de a considera problema zarafului „o bagatelă neserioasă": într-adevăr, această problemă a devenit un model pentru rezolvarea uneia dintre problemele practice importante, problema găsirii unor defecte în mașini complicate. E drept, problema rezolvată este foarte simplă, dar procedura utilizată aici, unirea variantelor în grupe și exami-

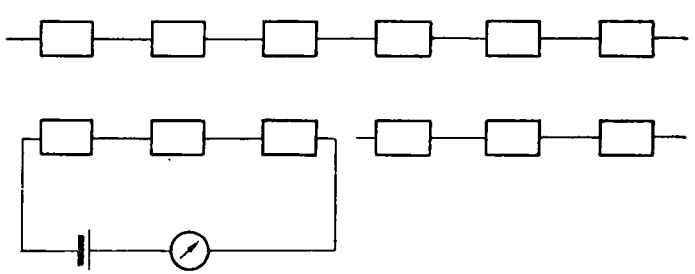


Fig 13.2

narea situațiilor posibile, este suficient de generală și de eficace *.

Să observăm că problemele reale de căutare a deficiențelor se complică cel puțin din cauza a două noi împrejurări. Raționamentul nostru se ocupa cu căutarea unei deficiențe într-un grup de elemente (respectiv monede) egale. *Dacă însă elementele sînt diferite, atunci în unele deteriorările apar mai frecvent, iar în altele mai rar.* Specialistul are un grad de neîncredere diferit față de diversele porțiuni ale unei mașini complicate, alcătuită dintr-o mulțime de elemente și structuri. El poate totdeauna să indice pe cele mai „prielnice" pentru defecte și tocmai pe acestea le verifică în primul rînd. *Acest sistem de preferințe, mai precis de „suspiciuni", poate fi precizat matematic sub aspectul unui sistem de probabilități ca un anumit element să se afle într-o stare de nefuncționare, urmînd să ținem seama de acest fapt nou în organizarea căutării.* De exemplu, atunci cînd „încercăm" lanțul este indicat să demarcăm nu porțiuni aproximativ egale, ci grupe de elemente pentru care sînt aproximativ egale probabilitățile de apariție a defecțiunilor **.

În sfîrșit, trebuie să ținem seama și de împrejurarea că verificările sînt de mai multe feluri și că nici o verificare nu seamănă cu alta: elementele sînt diferite și cheltuielile legate de verificarea unora pot diferi mult de cheltuielile necesare pentru verificarea altora.

Ținînd seama de diferența dintre probabilitățile de cădere (avarie) și de costul diferit al verificării, devine cu totul nebanală chiar și stabilirea ordinii de succesiune a verificărilor în cazul cînd diferitele elemente nu sînt

---

* Cititorul care știe ceva electrotehnică înțelege desigur că metoda căutării monedei false cu ajutorul unui cîntar cu talere poate fi transpusă la căutarea unei întreruperi într-un circuit, dacă avem la dispoziție o *punte a lui Wheatstone*, între bornele căreia se pot include dintr-o dată cîte k elemente (fig. 13.3), deși, firește, și o asemenea „tehnică" nu poate decît să provoace zîmbetul unui specialist.

** Cititorul care posedă o pregătire în domeniul teoriei probabilităților și este interesat de problema verificărilor optime poate găsi un material interesant în lucrarea [12].

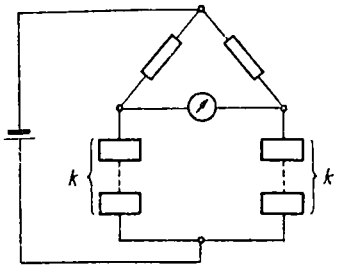


Fig. 13.3

reunite în grupe. De altfel, ideea de bază este limpede: din întreaga succesiune de elemente trebuie verificat în primul rînd acel element la care probabilitatea de cădere este mai mare, iar costul verificării mai mic. Da, dar dacă verificarea elementului celui mai capricios este cea mai scumpă? Atunci de bună seamă că trebuie să ne conducem după raportul dintre costul verificării și probabilitatea de cădere. Un raționament mai riguros arată că în primul rînd trebuie verificat elementul pentru care este minimă mărimea $\frac{cq}{p}$, unde c este costul verificării, p este probabilitatea de cădere a elementului respectiv, iar $q = 1 - p$ probabilitatea absenței avariei.

Ca exemplu, se dă un tabel cu caracteristicile a patru elemente.

N	c	p	q	$\frac{cq}{p}$
1	10	0,1	0,9	90
2	20	0,8	0,2	5
3	15	0,4	0,6	22,5
4	30	0,6	0,4	20

La început trebuie să fie verificat elementul 2. Dacă el funcționează, se verifică apoi 4, 3 și la sfîrșit 1. Îndată

ce se descoperă elementul defect, nu mai este necesar să se verifice celelalte *. Depinde de întîmplare dacă va trebui să se verifice un singur element, două sau mai multe, dar întreaga procedură este alcătuită astfel încît verificarea să aibă loc cu minimizarea cheltuielilor medii (după toate variantele posibile, adică în terminologia introdusă în cap. 8, cu minimizarea speranței matematice a cheltuielilor).

Cititorul poate că-și mai amintește cum încă la începutul cărții s-a relatat trista istorie a defectării unui calculator.

Acum putem privi de pe alte poziții munca inginerilor care caută defectele mașinii cu ajutorul unor programe-test speciale. Într-adevăr, o mașină modernă de calcul cuprinde un număr uriaș de elemente. Este de neconceput ca ele să fie verificate cîte unul; tocmai aici funcționează principiul reunirii în grupe.

Mașina este pusă să execute un calcul în cursul căruia este solicitat un mare număr de elemente. Rezultatul corect al calculului este cunoscut în prealabil. Dacă rezultatul oferit de mașină este altul, e clar că unul dintre elementele participante la calcul este avariat. Dacă însă mașina a rezolvat problema în conformitate cu răspunsul, înseamnă că toți participanții la calcul sînt în ordine și trebuie verificate numai elementele rămase.

Și aici, aceeași problemă a „monedei false"! Dacă toate elementele ar fi egal suspectate de defectare, atunci și problemele de verificare ar trebui construite astfel încît în cursul rezolvării să fie antrenate aproximativ o jumătate din elementele neverificate. Cum anume se nășcocesc asemenea probleme de verificare, probleme-test, aceasta este iarăși o problemă, dar examinarea ei ne-ar îndepărta prea mult de drumul pe care ni l-am ales.

Dealtfel, ar fi timpul să ne dăm seama dacă acest drum este într-adevăr atît de bun, dacă nu există altele mai bune

---

* Se presupune că probabilitatea de *cădere simultană* a două elemente sau chiar a unui număr mai mare este neglijabil de mică.

și totodată să lămurim în genere cum trebuie ales drumul cel mai bun. Dar, pentru început, un „mini-roman" detectiv*.

...Primele luciri ale zorilor nu biruiseră încă văpaia luminilor din Jazz-City, cînd sergentul McQuashkee, care își făcea rondul pe șoseaua interurbană, descoperi în ceața matinală ceva suspect. Se apropie în grabă de locul care-i atrăsese atenția și se încredință că nu se înșelase: în șanț se afla un corp, în mod evident neînsuflețit. Dealtfel, acesta era singurul lucru neîndoielnic, în rest totul era dubios în cea mai mare măsură. Pe fața tînărului mort înghețase un zîmbet, iar în palma dreaptă semideschisă se afla un revolver. Părea să fie vorba de o sinucidere: tînărul și palidul intelectual se hotărîse să pună capăt unei vieți care-i devenise nesuferită și se bucura cu anticipație că va scăpa de toate poverile ei. Și totuși: ce nevoie avea un tînăr intelectual să-și pună capăt vieții tocmai la marginea unei șosele interurbane? În buzunarele mortului nu s-au găsit nici bani, nici acte, dar, judecînd după costumul cumpărat pe Bond Street, dacă defunctul avusese necazuri, nu lipsa de bani fusese cauza lor. Și totuși sergentul reuși să găsească în buzunarul hainei ceva interesant. un lung fir de păr auriu. „Va trebui căutată femeia" — mormăi bătrînul McQuashkee.

*Cadavrul, vestonul și firul de păr fură aduse la sediul poliției, unde experimentatul criminalist D.T. Cteve se apucă de îndată de cercetări. Microanaliza structurii firului de păr dovedi convingător că posesoarea sa era o anumită Jane Pshephlee**, studentă la Școala Superioară de Afaceri. Polițiștii, în frunte cu D.T. Cteve, se duseră îndată la locuința ei, dar, vai! proprietara nu se mai afla acolo. Percheziția nu dădu nici un rezultat, cu excepția descoperirii unei foi de hîrtie uitate pe masă. Pe hîrtie era ceva ca o schemă și un tabel cu numere (fig. 13.4).*

Unul din bravii polițiști fusese cît pe ce să arunce această fițuică de pe masă, dar D.T. Cteve n-ar fi fost detectiv dacă

---

* Acest mini-roman a fost inclus în text pentru distracția cititorilor neserioși, dar conținutul lui matematic este mai mult decît serios.

** Autorul a dat eroilor acestei istorioare nume-rebus. *D.T. Cteve* = detectiv; *she flee* = ea fuge; *quash key* = îngroapă-soluție (l. engleză) (*N.T.*).

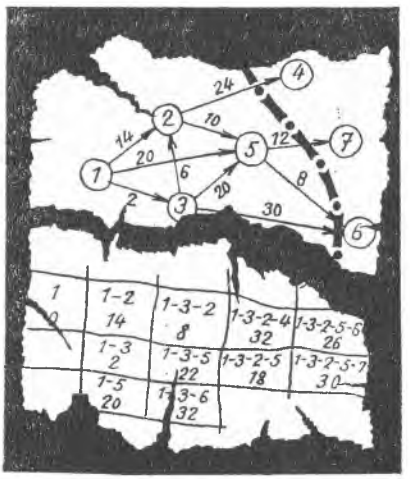


Fig. 13.4

hîrtia nu i-ar fi atras atenția: prea se vedea că notațiile erau cu totul recente. Analiza cernelii, efectuată în laborator, confirmă în mod convingător că nu se putea presupune că în clipele în care se pregătea să fugă domnișoara Pshephlee nu avusese de făcut nimic mai grabnic decît să-și continue obișnuitele ei preocupări studențești.

Misterioasele calcule aveau evident o legătură cu crima, dar care anume? — iată ideea care i se tot învîrtea în capul lui D.T. Cteve. Iar și iar relua detectivul studiul notațiilor, dar în zadar. În sfîrșit, îl străfulgeră o idee cu adevărat genială: „Să ne adresăm matematicienilor!" Fițuica fu expediată la Centrul de calcul al firmei YMOD, ceea ce înseamnă: „Voi cu banii, noi cu calculele".

Să nu așteptăm rezultatele anchetei și să încercăm să ne descurcăm singuri. Logica lucrurilor este cît se poate de limpede: dacă a avut loc o crimă, criminalul trebuia să fi fugit cît mai repede. Încotro? Oriunde dincolo de granițele iubitei sale patrii, dar urmînd drumul cel mai bun!

Așadar, presupusa asasină avusese de rezolvat o problemă de alegere a drumului optim. Ca studentă la Școala

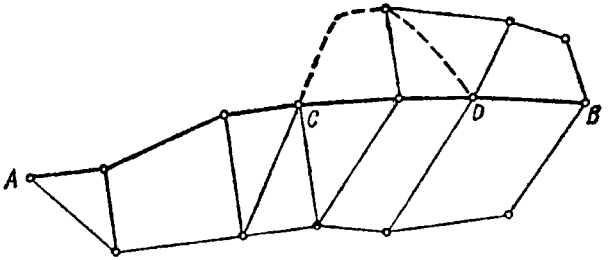


Fig. 13.5

Superioară de Afaceri, era, desigur, în stare s-o rezolve cu ajutorul matematicii, prin metodele cele mai moderne. Acum totul este clar: schema reprezintă harta simplificată a drumurilor ce duc la frontieră, iar numerele — timpul necesar pentru a trece de la un punct nodal la altul, punctele nodale fiind notate de la 1 la 7.

Să admitem. Dar mai departe?

Mai departe va fi nevoie să facem cunoștință cu principiile celei mai moderne metode de căutare a drumului celui mai bun, metodă propusă de savanții americani R. Bellman și G. Dantzig *. Din fericire, aceste principii sînt foarte simple.

Să presupunem că drumul cel mai bun din punctul A în punctul B trece prin punctele C și D (linia groasă din fig.13.5). Să presupunem acum că condițiile s-au schimbat și se cere doar să se treacă din C în D. Care va fi drumul optim din C în D? Este evident că el trebuie să coincidă cu o porțiune a drumului optim din A în B. Într-adevăr, dacă acest lucru n-ar fi adevărat și dacă ar fi mai bine să se pornească dinspre C pe o altă cale (de exemplu pe calea indicată prin linie punctată în fig.13.5), atunci și drumul optim de la A la B n-ar fi cel dat: ar fi mai avantajos la început să parvenim din A în C pe drumul dat, apoi să pornim înspre D pe drumul punctat; astfel cheltuielile ar fi mai mici. Dar, așa cum am afirmat de la început,

* De astă dată numele sînt reale și realmente renumite [2, 3].

drumul dat de la *A* la *B* este optim, cel mai bun dintre toate drumurile accesibile, deci un cîștig oarecare în raport cu acesta este imposibil.

Așadar, drumul optim între două puncte oarecare situate pe traseul optim dintre alte două puncte trebuie să fie o parte a acestui traseu, sau mai pe scurt: **orice porțiune a unui drum optim este de asemenea drum optim.**

Tocmai acesta este renumitul *principiu al optimului* propus de către R. Bellman. Utilitatea unui asemenea principiu este evidentă: găsind drumul cel mai bun între două puncte intermediare, putem fi pe deplin încredințați că *dacă unul din aceste puncte se află în drumul nostru, etapa următoare trebuie efectuată și ea după drumul optim găsit, fără a mai lua în considerare vreo altă variantă.* Necazul constă doar în faptul că nu putem spune dinainte dacă aceste puncte ne vor fi sau nu în drum. Este clar că punctul inițial se va afla cu siguranță pe drumul nostru, celelalte le vom „bănui" rînd pe rînd că ar fi optime.

Acum este însă timpul să revenim la problema lui D.T. Cteve și să vedem cum se realiza ideea generală a căutării drumului optim în planul concret al crimei. La început să privim schema. Din Jazz-City, numerotat cu 1, putem trece de îndată (fără a trece prin vreun alt punct) în punctele 2, 3 și 5. Orice drum către frontieră dinspre 1 va conține drept secțiune inițială fie 1 — 2, fie 1 — 3, fie 1 — 5. Să presupunem că inițial este drumul 1 — 2. Atunci drumul poate fi prelungit pînă la 3, 4 sau 5 (deoarece, evident, a reveni la 1 nu este indicat).

Analog, după 1 — 3 pot urma sectoarele 3 — 2, 3 — 5 sau 3 — 6, iar după 1 — 5 pot urma sectoarele 5 — 3, 5 — 6 sau 5 — 7. Fiecare dintre aceste drumuri „cu două sectoare" (de obicei se spune „în doi pași"), cu excepția celor care se încheie cu puncte situate pe frontieră, poate fi prelungit cu cel puțin un sector; și așa mai departe. Drept urmare, vom obține schema tuturor variantelor posibile prezentată în figura 13.6. O asemenea schemă se numește de obicei

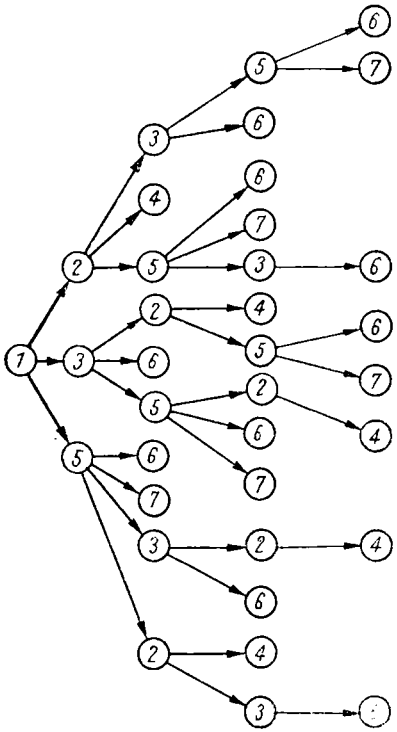


Fig. 13.6

arborele variantelor *. *La fiecare extremitate a numeroaselor sale ramuri se află fructul rîvnit, un punct de frontieră, iar numărul extremităților este 20!*

Pentru a alege calea cea mai bună, se poate utiliza procedura „grosieră" a trierii directe, socotind timpul necesar pentru parcurgerea fiecărei variante în parte. Dar merită oare? Avem doar la dispoziție principiul lui Bellman și

* Această schemă amintește schema rezultatelor posibile cînd se aruncă cu banul de mai multe ori (cap. 8), numai că aici alegerea nu se realizează la întîmplare, ci prin estimarea calității variantelor.

el ne va îngădui să renunțăm la numeroase variante inutile fără a le mai examina pînă la capăt.

Să calculăm consumul de timp numai după variantele pașilor inițiali. Dealtfel, ele se pot transcrie direct de pe „harta drumurilor". Oricum ar continua drumurile ulterioare, timpul consumat de un sector inițial n-are cum să se schimbe. Să-l notăm pe schemă alături de numărul de ordine al punctului (fig. 13.7 a). Să facem aceeași operație cu toate variantele drumurilor în doi pași, așezînd alături de numărul de ordine al punctului atins timpul consumat în total de cele două sectoare. În continuare am putea să examinăm și toate drumurile în trei pași, însă tocmai acum e timpul să chemăm în ajutor principiul optimului! Să presupunem că punctul 2 se află pe drumul optim spre frontieră; numai în acest caz merită să ne interesăm de drumurile care conduc la 2. Dar în acest caz trebuie să înlăturăm toate drumurile de la 1 la 2 cu excepția celui mai bun, a celui mai rapid, drumul 1 — 3 — 2, care pretinde în total 2 + 6 = 8 ore. Amintim iarăși că o porțiune a drumului optim, oricare ar fi ea, trebuie să fie de asemenea drum optim („Noblesse oblige" — noblețea îți impune îndatoriri — după cum se spunea în „bunele vremuri de demult"!) Dar prin aceasta obținem posibilitatea să „retezăm din rădăcină" toate variantele (iar ele sînt nu mai puțin de șapte!) începînd cu sectorul inițial 1 — 2, care pretinde 14 ore, precum și alte două variante, prelungiri posibile ale drumului 1 — 5 — 2, care pretinde 20 + 10 = 30 ore. Folosind același raționament referitor la drumurile care conduc către punctele 3 și 5, vom înlătura alte cinci variante nefericite. În plus, comparînd drumurile care conduc la frontieră, vom menține numai pe cel mai bun, drumul 1 — 5 — 6, care pretinde 20 + 8 = 28 ore, ca fiind unicul bănuit de a fi optim.

În urma acestor considerente, totul s-a modificat în mod hotărît: dintre drumurile în trei pași posibile merită studiate numai două variante (fig. 13.7 c), reținînd în memorie și rezultatul unui drum în doi pași. Aici iarăși apare posibilitatea de a înlătura o variantă încheiată (1 — 3 — 2 — 4) ca fiind mai proastă decît cea găsită, iar apoi înlăturăm și

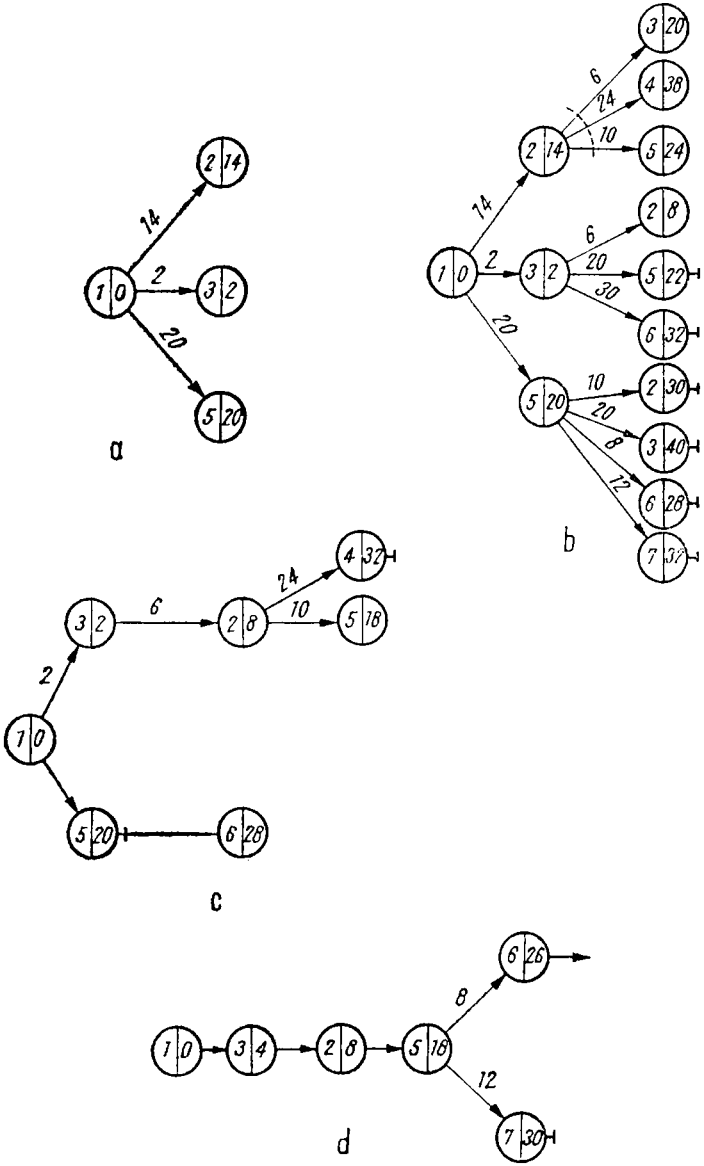

a
1 0
2 14
3 2
5 20
14
2
20
b
1 0
2 14
3 2
5 20
3 20
4 38
5 24
2 8
5 22
6 32
2 30
3 40
6 28
7 32
14
2
20
6
24
10
6
20
30
10
20
8
12
c
1 0
3 2
2 8
5 20
4 32
5 18
6 28
2
6
24
10
d
1 0
3 4
2 8
5 18
6 26
7 30
8
12

Fig. 13.7

*pe cea găsită (1 — 5 — 6), deoarece conține sectorul 1 — 5, vădit mai prost decît drumul 1 — 3 — 2 — 5. În sfîrșit, comparația efectuată la pasul al patrulea (fig. 13.7 d) scoate la iveală unicul * drum optim: 1 — 3 — 2 — 5 — 6 care consumă doar 26 de ore. Nu-i așa că această metodă este cu mult mai simplă decît trierea directă?* Ca și metoda zarafului, această metodă utilizează examinarea unor grupe întregi de variante și tăierea grupelor inadecvate. E adevărat că variantele nu se pun pe cîntar, ci se estimează doar după porțiunea inițială comună, dar aceasta nu mai este o deosebire de principiu.

Firește, acum putem comunica poliției din Jazz-City drumul urmat de criminala Pshephlee și astfel să-l izbăvim de tortura intelectuală pe semidoctul în matematică D.T.Cteve. Dar o siguranță deplină încă nu avem, deoarece notele găsite în locuința criminalei nu seamănă cu figurile cu arbori de care am făcut uz. Dar în această problemă ne putem descurca ușor. Cititorul desigur că și-a dat seama că pentru rezolvarea problemei nu era necesar să se figureze întregul arbore al variantelor posibile. Este suficient să desenăm arborele numai în procesul creșterii, retezînd de la început ramurile care cu siguranță nu promit roadele optime. Dealtfel și desenul este un lux exagerat, un tribut plătit concretului. Calculul poate fi așezat în formă de tabel.

Executînd tabelul, obținem ceva asemănător cu notele lui Jane Pshephlee, dar nu întru totul: tabelul ei conține mai puține calcule.

Examinîndu-l cu atenție, putem constata încă o idee simplă. De ce să construim dintr-o dată toate variantele într-un pas, apoi toate în doi pași și așa mai departe? Nu-i mai bine ca în calcul *să dezvoltăm numai drumurile cele mai „bănuite de optim" pornind din punctul inițial?*

Să aplicăm această idee în mod consecvent. Ca și înainte, să examinăm toate drumurile de un pas care pornesc dinspre 1.

---

* Faptul că drumul optim este în cazul de față unic nu e o regulă; sînt posibile și probleme unde apar mai multe drumuri egale în ce privește cheltuielile totale și, ca atare, optime.

1 0	1—2 14	1—2—3 20	1—3—2—4 32	1—3—2—5—6 26
	1—3 2	1—2—4 38	1—3—2—5 18	1—3—2—5—7 30
	1—5 20	1—2—5 12		
		1—3—2 8		
		1—3—5 22		
		1—3—6 32		
		1—5—2 30		
		1—5—3 40		
		1—5—6 28		
		1—5—8 32		

*Ne lămurim că cel mai bun, prin urmare și cel mai bănuit de a fi începutul drumului optim este sectorul 1 — 3, cu un consum de 2 ore. Să desfășurăm tocmai această variantă, studiind rezultatele adăugirii la ea a încă unui sector și înghețînd deocamdată pe celelalte. Să comparăm acum toate cele cinci variante studiate (restul celor într-un pas și cele în doi pași „crescute" din 1 — 3). Găsim că cea mai eficientă dintre ele este varianta în doi pași 1 — 3 — 2 și în continuare o vom dezvolta cu încă un pas *, înlăturînd totodată*

* Dacă mai bună s-ar fi dovedit acum una din variantele de un pas, am dezvolta-o, completînd astfel lista celor în doi pași.

și varianta 1 — 2, vădit inaptă de a fi dezvoltată în ceva mai bun decît 1 — 3 — 2. Obținem variante noi, dar acum în trei pași: 1 — 3 — 2 — 4 și 1 — 3 — 2 — 5. Selecția din cele patru drumuri existente și apte de a fi dezvoltate promovează ca cel mai bun drumul 1 — 3 — 2 — 5 și dă la o parte 1 — 5 și 1 — 3 — 5 ca fiind cu totul inapte pentru continuare. Ultimul pas conduce exclusiv la variante care se încheie pe frontieră și printre acestea există una mai bună decît toate de care dispunem *.

Ajungem astfel definitiv la varianta inscripției care a apărut în povestirea polițistă. Cu aceasta se poate considera ca fiind pe deplin încheiată afacerea crimei.

Succesul constă, firește, nu în faptul că a apărut posibilitatea de a o prinde pe criminală, ci în acela că am analizat principiul optimului și utilizarea acestuia ca instrument care permite retezarea ramurilor inutile din arborele variantelor și totodată principiul auxiliar propus de G. Dantzig, care simplifică și mai mult lucrurile.

De faptul că arborele variantelor înclină spre o creștere amenințătoare ne putem lesne convinge prin exemple simple. Să presupunem că schema drumurilor posibile către frontieră are un aspect simetric, indicat în figura 13.8, și că sînt admise numai drumurile fără reveniri. Atunci la fiecare pas numărul variantelor se dublează. Dacă numărul punctelor nodale așezate pe catete este N, numărul variantelor de sosire în punctele ipotenuzei (pornind din vîrful 1) este 2^N. Cititorul probabil că-și amintește că trierea directă a 2^{100} variante este o sarcină fără speranțe. Utilizînd însă principiul optimului, reușim destul de ușor să facem față acestei probleme.

Dar poate că nu merită să facem eforturi pentru că domeniul de aplicare a unor asemenea metode n-ar fi decît rezolvarea problemelor puse de unele povestiri elementare cu detectivi? În realitate, lucrurile nu stau deloc astfel. Problema căutării drumului celui mai bun are o foarte

---

* Trebuie comparate toate variantele păstrate, inclusiv cele care nu asigură deocamdată ieșirea la frontieră. Cu totul întîmplător, în problema de față n-au apărut asemenea variante.

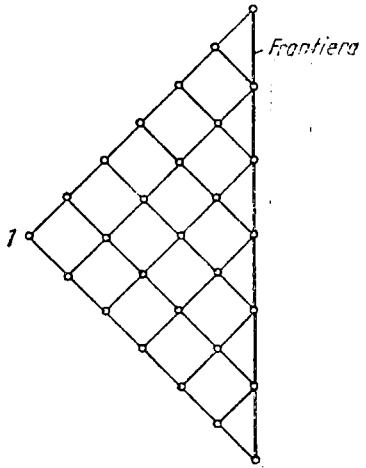


Fig. 13.8

mare importanță în cele mai diferite aplicații, nu numai în transporturi, ci și în crearea sistemelor de conducere a avioanelor, rachetelor și sateliților.

Prima problemă a unui avion de vînătoare este să ia înălțime și viteză cît mai rapid posibil.

Acești doi factori determină în mare măsură posibilitățile de luptă ale mașinii. Superioritatea în înălțime și viteză este de regulă chezășia succesului în atac. De aceea, luîndu-și startul în întîmpinarea luptei, aviatorul tinde să atingă înălțimea cerută — suficient de mare — într-un timp minim, și anume cu o viteză dată foarte mare. Pentru avioanele de luptă moderne, înălțimea este de 20—30 km, adică stratosfera, iar vitezele depășesc de 2—3 ori viteza sunetului, adică sînt de 2 000 — 3 000 km pe oră. Problema de a alege drumul, adică traiectoria care te conduce cel mai rapid la regimul fixat este indubitabil o problemă importantă și, tot atît de indubitabil, ea este o problemă de căutare a drumului optim, problemă a cărei variantă simplistă a fost descrisă în povestea cu asasinatul.

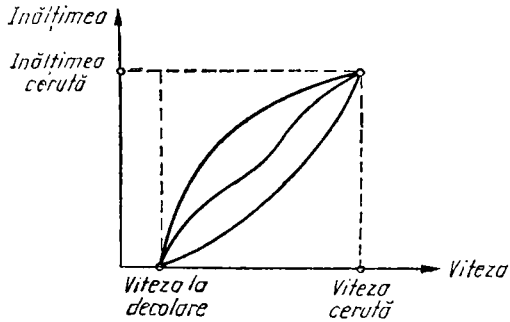


Fig. 13.9

Toate drumurile posibile pot fi reprezentate ca niște curbe în planul cu axele de coordonate „viteză-înălțime", curbe care unesc punctul startului și punctul final corespunzător valorilor date pentru viteză și înălțime (fig.13.9).

Pilotul poate alege curbele (traiectoriile) cele mai variate. În principiu, în fiecare moment, cu ajutorul dispozitivelor de conducere ale avionului, el poate să dea cele mai variate valori unghiului de înclinare a traiectoriei față de suprafața pămîntului și prin urmare să modifice aspectul întregului drum rămas de parcurs.

Funcția-scop, timpul de zbor pe traiectorie, este astfel dependentă de o infinitate de variabile: valorile unghiului de înclinare a traiectoriei la fiecare moment de timp.

Dificila problemă a căutării optimului unei asemenea funcții poate fi, desigur, simplificată considerînd că schimbările pot avea loc numai într-un număr finit de puncte ale traiectoriei. Cu cît vom lua mai multe asemenea puncte, cu atît mai mult ne vom apropia de rezolvarea problemei inițiale *. Să simplificăm problema noastră în două feluri. Mai întîi, să presupunem că modificările în comanda avio-

* Aici trebuie căutată una din principalele surse ale apariției problemelor de optimizare în spațiul multidimensional despre care s-a vorbit încă în cap. 3.

nului aviatorul le poate introduce numai în anumite puncte fixate din planul „viteză-înălțime", de exemplu punctele situate în nodurile rețelei 10×10, care descompune în mod uniform domeniul variației înălțimilor și vitezelor, de la cele inițiale pînă la cele finale (fig. 13.10). În al doilea rînd, vom examina numai traiectoriile care trec prin aceste puncte, și anume fără „reducerea" vitezei sau înălțimii obținute.

În felul acesta, numărul total de momente în care pilotul poate să modifice direcția zborului devine finit, mai exact: egal cu 20. De asemenea, alegerea variantelor în fiecare din aceste momente este simplificată la extrem: în fiecare punct (cu excepția celor de frontieră) se poate: 1) sau mări viteza menținînd constantă înălțimea sau 2) mări înălțimea menținînd constantă viteza. Dar și cu această simplificare numărul variantelor diferitelor traiectorii de la punctul startului înspre punctul finișului rămîne uriaș. Se poate calcula că el este aproximativ egal cu 185 000! Așadar, și aici posibilitatea trierii directe nu pare acceptabilă.

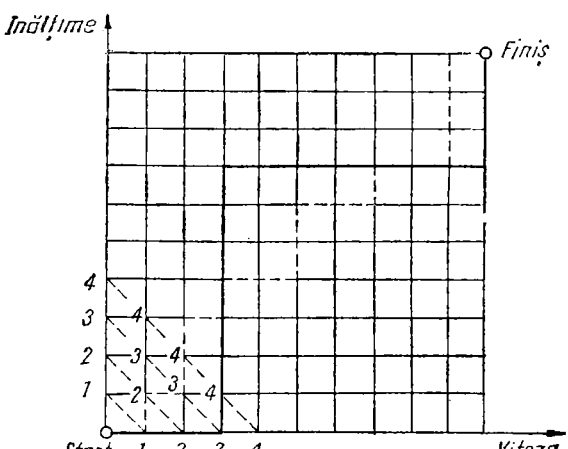


Fig. 13.10

Dar iată că problema a căpătat același aspect pe care l-a avut problema căutării drumului optim în „povestea" crimei! Numai că în locul posibilelor puncte intermediare pe drumul către frontieră figurează nodurile pe rețea, iar frontiera însăși se reduce la un singur punct dorit: finișul. Consumul de timp necesar pentru a trece prin fiecare din sectoarele intermediare ale drumului poate fi calculat după caracteristicile avionului, greutatea sa, forța de tracțiune, dependența rezistenței aerului și a forței ascensionale de înălțimea și viteza zborului.

Dacă asemenea calcule sînt efectuate de către specialiștii în aerodinamică, obținem o schemă totală a drumurilor posibile și putem folosi metoda cunoscută de tăiere succesivă a variantelor inadecvate. Este de remarcat că într-un anumit sens această problemă a obținerii înălțimii este mai simplă chiar decît cea „polițistă". În virtutea simetriei schemei și a ipotezei adoptate cu privire la inadmisibilitatea mișcărilor „în jos și la stînga", punctele schemei se descompun în mod natural în etaje.

În fiecare etaj (fig. 13.10) intră acele puncte la care se poate ajunge prin unul și același număr de pași (de exemplu 4 pași), etajele neputîndu-se intersecta: dacă într-un punct ajungem prin N pași, atunci nu este posibil să ajungem la el printr-un număr mai mic sau mai mare de pași *. Prin aceasta, în problema obținerii înălțimii se degajează cu ușurință mulțimile de variante cu un pas, cu doi pași și așa mai departe, ceea ce simplifică și ordonează calculele.

Să dăm unele formule pentru calculul drumului optim pe o rețea cu un număr N de sectoare după fiecare coordonată. Numărul total al sectoarelor (pașilor) pe orice traiectorie care conduce din punctul inițial în punctul final este $2N$. În mod corespunzător, toate nodurile rețelei se pot descompune în $2N + 1$ etaje, în fiecare intrînd nodurile accesibile în k pași ($k = 0,1, ..., 2N$). Să numerotăm de asemenea punctele din fiecare etaj, notînd cu t_{ij}^{k} timpul consumat în cursul unui pas din punctul al i-lea din etajul al k-lea către punctul al j-lea din etajul al $(k + 1)$-lea.

* În schema „polițistă" punctul numerotat cu 4 putea fi atins printr-un număr variabil de pași: 2, 3 sau 4.

Să definim $T_j(k)$ ca timpul minim necesar pentru a ajunge din origine în punctul al j-lea din etajul al k-lea. Atunci

$$T_j(0) = 0, \quad j = 1$$

(în etajul nul există un punct unic, tocmai punctul inițial, originea). În continuare, este clar că

$$T_j(1) = t^0_{1j}, \quad j = 1,2$$

(în fiecare al j-lea punct din primul etaj se poate nimeri printr-un singur drum de un pas din unicul punct al etajului nul, consumînd un timp cunoscut).

Dezvoltînd variantele în variante cu doi pași, cu trei pași etc., putem estima timpul $T_i(k)$ pentru orice punct al celui de-al k-lea etaj, inclusiv pentru punctul final, unicul din ultimul etaj ($k = 2N$). Dar aici, desigur, trebuie înlăturate variantele vădit proaste folosind principiul optimului. Ne putem convinge că pentru orice k este valabilă formula

$$T_j(k+1) = \min_i [t^k_{ij} + T_i(k)], \text{ unde } k = 0, 1, 2, \ldots, 2N-1,$$

minimul calculîndu-se prin trierea după toate punctele celui de-al k-lea etaj. Într-adevăr, drumul către punctul al j-lea al celui de-al $(k+1)$-lea etaj este alcătuit din $k+1$ pași. Dar în acest punct se poate ajunge numai dintr-un punct al etajului precedent.

Prin urmare, orice drum avînd $k+1$ pași trebuie să fie alcătuit din două părți: un sector de k pași care conduce din origine către un anumit punct al celui de-al k-lea etaj, să zicem punctul al i-lea, și un sector de un pas care constituie trecerea de la acest punct la punctul al j-lea din etajul al $(k+1)$-lea. Dar dacă acest drum este optim, atunci și prima sa parte trebuie să fie optimă și deci îi revine timpul $T_i(k)$.

Formula de mai sus reflectă numai necesitatea alegerii optime a porțiunii de un pas care a mai rămas de efectuat. Totodată, ea permite să se calculeze timpul optim de trecere în etajul al $(k+1)$-lea atunci cînd se cunoaște timpul optim de trecere în etajul al k-lea, și întrucît pentru $k = 1$ mărimile $T_j(k)$ sînt date, apare implicit posibilitatea de a ajunge succesiv la $T_j(2N)$, adică timpul optim al trecerii la punctul final cu înlăturarea variantelor inadecvate.

Drept exemplu, să considerăm calculul timpului optim de luare a înălțimii dacă problema este simplificată în prealabil cu ajutorul rețelei 3×3. În figura 13.11 este reprezentată această rețea, pentru fiecare sector fiind indicat timpul necesar pentru parcurgerea sa. În fiecare nod se indică numărul de ordine al etajului și numărul de ordine în interiorul etajului. Calculul se urmărește ușor după tabel.

	k=1		k=2		k=3		k=4		k=5		k=6	
j	i(j)	Tj(1)	i(j)	Tj(2)	i(j)	Tj(3)	i(j)	Tj(4)	i(j)	Tj(5)	i(j)	Tj(6)
1	1	60	1	160	1	280	2	240	2	260	2*	300
2	1*	40	2*	80	2*	140	2*	160	2*	240		
3			2	110	2	140	3	200				
4					3	190						

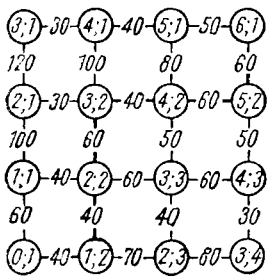


Fig. 13.11

Pentru fiecare etapă avînd numărul de ordine k, tabelul dă mărimea timpului optim de acces în fiecare al j-lea punct din etajul al k-lea, $T_j(k)$, precum și numărul de ordine al punctului din etajul precedent, $i(j)$, din care este cel mai convenabil să ne îndreptăm spre punctul dat.

Să examinăm mai amănunțit cum se realizează, de exemplu, etapa a 4-a. La început privim punctul întîi ($j = 1$) din etajul al 4-lea. Aici se poate ajunge prin două drumuri: a) sau din primul punct al etajului al 3-lea, consumînd pentru aceasta 30 s, ceea ce împreună cu timpul consumat anterior $T_1(3) = 280$ s va da 310 s; b) sau din punctul al 2-lea din etajul al 3-lea, consumînd pentru aceasta 100 s, dar aici consumul anterior de timp este $T_2(3) = 140$, astfel încît în acest al doilea caz consumul total de timp este $100 + 140 = 240$ s, ceea ce este mai puțin decît 280 s și, prin urmare, reprezintă timpul cel

mai redus posibil pentru a ajunge din origine în punctul $j = 1$ din etajul al 4-lea. De aceea se poate înscrie în tabel $T_1(4) = 240$, iar acest timp se obține dacă sosim în punctul dat din punctul cu numărul de ordine $i = 2$ al etajului precedent. Exact la fel (și anume cu mult mai rapid decît se poate relata prin vorbe) se face completarea celorlalte celule ale tabelului.

Timpul optim pentru a lua înălțime se constată a fi egal cu 300 s. Totodată se poate indica și drumul optim pe baza tabelului. Pentru aceasta este necesar să înaintăm pînă la capăt. Se constată că în punctul final ar fi cel mai util să ajungem din punctul al 2-lea al etajului al 5-lea. La rîndul său, acest punct este accesibil cu consum minim de timp din punctul al 2-lea din etajul al 4-lea. Și așa mai departe. Mergînd înapoi și utilizînd datele coloanelor unde sînt așezate numerele de ordine $i(j)$, aflăm toate punctele traiectoriei optime (în fig. 13.11 ea este marcată prin linia groasă, iar punctele corespunzătoare sînt marcate cu stea). Aceasta-i toată tehnica calculului * !

Acum să ne ocupăm iarăși de „ideologia" calculului. Cititorul atent probabil că a observat că în exemplul nostru traiectoria optimă se putea obține și pe altă cale mult mai simplă. Să rezolvăm problema alegerii drumului optim prin metoda cea mai „mioapă", adică fără a ne interesa de caracterul optim al întregii traiectorii, ci numai de consumul de timp pretins de sectorul cel mai apropiat al drumului. Din punctul (0; 1) pornesc două drumuri. Să alegem drumul care cere un timp mai redus; parvenim astfel în punctul (1; 2). Aici iarăși este de ales din două drumuri posibile; aplicînd principiul „avantajului proxim",

---

* Întrucît în acest exemplu numerele, deși convenționale, aveau totuși mai mult sau mai puțin sens, rezultatul obținut seamănă și el cu cel real: avionul merge la început pe o traiectorie puțin înclinată, mărindu-și viteza. Apoi, înainte de a ajunge la viteza sunetului, ia brusc înălțime. Luînd înălțime, își ia avînt, străbătînd bariera sonoră. Apoi continuă iarăși să ia înălțime pînă ajunge la cea determinată în prealabil. În lucrarea [3] se poate găsi o descriere mai amănunțită a metodei de rezolvare a acestei probleme cu ajutorul principiului optimului.

ajungem în (2; 2). Din acest punct pornesc două drumuri care pretind timpuri egale. Pentru a alege unul din ele, să privim cu doi pași înainte. Vom ajunge prin (3; 2) în (4; 2), iar de acolo aceeași „miopie" ne va aduce la țelul final fără nici un fel de complicații. Și, oricît de ciudat ar părea, se va constata că am ales tocmai acel drum care a fost recunoscut drept cel mai bun dintre toate drumurile posibile.

Mai mult; se constată că o asemenea coincidență nu este întîmplătoare, ci caracteristică pentru multe probleme de traiectorie optimă a zborului, deși, desigur, nu pentru toate aceste probleme. Această plăcută împrejurare ne ușurează mult problema optimizării. Chiar dacă traiectoria „miop-optimă" (în știință ea este denumită *cvasistatic optimă*) nu este tocmai cea mai bună, ea poate fi adoptată în primă aproximație, pentru a încerca apoi îmbunătățirea prin aprecierea valorilor funcției-scop pe traiectorii învecinate, care se deosebesc de această traiectorie numai pe sectoare mici.

În figura 13.12 *a* se indică schema drumurilor care unesc punctele *A* și *B*. Dacă ne conducem după o politică „mioapă", vom alege drumul indicat prin linia groasă care cere 17 ore. Putem să încercăm să-l îmbunătățim înlocuind diferite sectoare ale drumului cu altele apropiate. Aproape toate drumurile modificate se vor dovedi mai proaste. Reușită va fi doar înlocuirea a două sectoare (care pretind un drum total de 8 ore) printr-unul care necesită numai 5 ore, astfel încît drumul total de la *A* la *B* va ocupa doar 14 ore. Modificările locale nu oferă nimic altceva. Totuși, nu e greu de observat că din *A* în *B* putem ajunge și în curs de 12 ore (fig. 13.12 *b*). Iată că am pierdut din vedere acest drum realmente optim numai din cauza „miopiei" de un pas. Dacă am fi comparat cel puțin combinațiile de doi pași, atunci el ar fi sărit dintr-o dată în ochi. Dar cu cît combinațiile sînt mai lungi, cu atît este mai dificil să le trecem în revistă pe toate; treaba

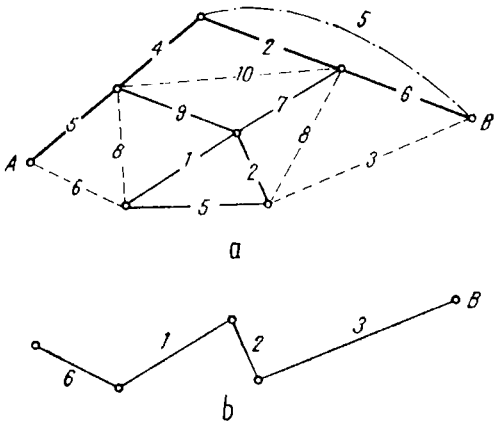


Fig. 13.12

aceasta va pretinde din ce în ce mai multă muncă, care în ultimă instanță se poate dovedi inutilă.

Probabil că cititorul a observat că am început să folosim un limbaj apropiat de cel al șahului. A trece în revistă rezultatele tuturor combinațiilor posibile de mai mulți pași (mai multe mutări) este practic imposibil; dacă însă ne limităm la combinațiile scurte, „mioape" (așa cum se procedează de obicei în multe programe primitive pentru participarea calculatorului la jocul de șah), pierdem de obicei perspectiva corectă.

Se pare că perfecționarea modului de alegere a drumului optim introdusă de G. Dantzig — dezvoltarea cu prioritate a variantelor „bănuite de a fi optime" — este pe deplin logică din punctul de vedere al jocului de șah. Autorul însă, fiind un șahist foarte prost, nu se poate decide să dea sfaturi experimentaților autori de programe pentru jocul de șah.

Firește, problemele de șah se deosebesc mult de problema alegerii drumului optim și prin aceea că în șah este

greu de apreciat utilitatea fiecărei mutări în parte; dar mai ales prin faptul că șahul este un joc la care participă doi parteneri urmărind țeluri direct opuse.

E adevărat că în neostoita noastră luptă cu Natura, care ne vîră cu tot dinadinsul variante proaste, noi am învățat într-o oarecare măsură să acționăm făcînd uz de *principiul asigurării cîștigului garantat.* Tocmai acest principiu se află la baza teoriei matematice a jocurilor, disciplină care utilizează multe idei generale din teoria căutării și este deosebit de interesantă *.

* Avînd în vedere această situație, ar fi firesc să examinăm acum problema alegerii deciziilor optime tocmai în cadrul jocurilor în care participanții au interese opuse. Dar volumul acestei cărți este mărginit, iar despre jocuri s-a vorbit destul în multe cărți foarte bune și chiar foarte simple, de exemplu [14, 29, 23, 27], la care poate face apel cititorul mînat de curiozitate științifică.

14
Secrete de producție

(Metoda celor mai mici pătrate)

Ei bine, asta este: prea ades
Vedem doar încotro pornește fumul,
Uitînd cu totul
Să ne întrebăm
De unde bate vîntul.

K. Čapek, *Despre lucruri*

Ca orice meserie onorabilă și cu vechime, căutarea își are secretele ei interne, de atelier, care determină succesul, dar rămîn neobservate de cei care se bucură doar de roadele muncii, fără a încerca să intre în bucătărie.

Se pare că principalul secret de producție al căutării, pe care l-am păstrat pentru cititorii care citesc cartea pînă la sfîrșit, îl constituie metodele de aproximare a funcțiilor.

Problema aproximării funcțiilor este absolut unică în istoria științei. Ea a avut o influență decisivă asupra celor mai abstracte și fundamentale idei din întreaga matematică. *S-au ocupat de această problemă aproape toți marii matematicieni din trecut, Newton și Lagrange, Gauss și Weierstrass, Cebîșev și Abel, ca și mari matematicieni contemporani cum sînt Kolmogorov și Wiener, Bernstein și Kantorovici. Pe de altă parte, nici una din problemele de matematică pură nu a fost atît de strîns și bilateral legată de practică, de la primele ei începuturi și pînă în prezent.*

În ce constă așadar problema aproximării funcțiilor? Se dă o funcție printr-o formulă, printr-un grafic sau prin valorile ei în diferite puncte. Se cere să se găsească o altă funcție, dintr-o familie dată de funcții, de exemplu un

polinom, care să se deosebească cît mai puțin de cea dată, să fie cît mai asemănătoare cu ea.

Din punctul de vedere al matematicii pure, abstracte, aceasta este problema deslușirii posibilităților de a reprezenta funcții dintr-o clasă prin funcții din altă clasă, prin urmare problema clarificării deosebirilor dintre clase.

Din punctul de vedere al matematicii aplicate, problema aproximării este problema găsirii unei metode eficiente de a descrie funcții mai complicate prin altele mai simple.

În sfîrșit, dacă ne situăm pe pozițiile tehnicii și a științelor naturii, ne întîlnim cu numeroase și variate probleme, vechi și noi, de la prelucrarea datelor experimentale pînă la recunoașterea tipurilor.

Să considerăm cel mai simplu circuit electric (fig. 14.1). Variind tensiunea aplicată circuitului, variem și curentul care curge prin el. Tensiunea o vom măsura cu voltmetrul, iar intensitatea curentului, după cum se și cuvine, cu ampermetrul. Orice elev cunoaște legea lui Ohm: tensiunea u este proporțională cu intensitatea curentului i, iar coeficientul de proporționalitate se numește rezistența R, adică $u = Ri$ sau invers $i = \rho u$, unde $\rho = 1/R$ este conductibilitatea. Altfel spus, legea lui Ohm afirmă că intensitatea curentului din circuit este funcție de tensiune, și anume o funcție liniară care este pe deplin determinată cînd se cunoaște o singură mărime: conductibilitatea.

Țelul experimentului nostru constă în determinarea acestei necunoscute, conductibilitatea circuitului. Pentru aceasta, în principiu este suficient să se măsoare intensitatea curentului corespunzătoare unei valori a tensiunii. De exemplu, să presupunem că sursa dă o tensiune de 15 V,

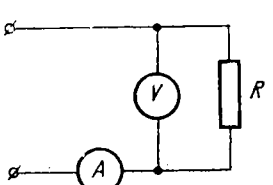


Fig. 14.1

iar intensitatea măsurată este de 0,6 A. Atunci calculăm fără dificultate că

$$R = \frac{15}{0{,}6} = 25 \text{ ohmi, adică } \rho = 0{,}04 \text{ 1/ohm.}$$

Să considerăm un alt exemplu clasic. Să observăm căderea unui corp greu într-un vas bine vidat (aproape vid). Dacă corpul cade fără viteză inițială (de exemplu, datorită desfacerii unui șurub), în timpul t el va trebui să cadă pe distanța

$$H = \frac{1}{2} gt^2 \text{ cm,}$$

unde g este accelerația constantă a căderii libere, a cărei cauză este atracția corpului de către pământ.

Legea mișcării cu accelerație constantă exprimată de această formulă spune că drumul parcurs de corp este funcție de timpul consumat, și anume o funcție pătratică determinată pe deplin prin cunoașterea unei singure mărimi: accelerația g. Faptul acesta vă este desigur bine cunoscut. Mai mult, probabil că vă este cunoscută și mărimea lui g, aproximativ egală cu 981 centimetri pe secundă (cm/s^2). Da, dar tocmai aici este buclucul: *aproximativ* egală. În realitate, mărimea g este diferită în diferitele puncte ale globului terestru, variind aproximativ cu $\pm$ 3 cm/s^2. Or, a cunoaște aceste variații este esențial și pentru prognoza exactă a mișcării sateliților și pentru treburi pur pământești. Dacă în fiecare punct de pe pământ g este diferit, atunci, alcătuind în prealabil harta distribuției acestor mărimi după rezultatele măsurătorilor exacte, putem în primul rînd să ne orientăm după aceste hărți, iar în al doilea rînd putem presupune că variațiile esențiale ale lui g indică modificări mari în distribuția mineralelor, ceea ce poate însemna o sugestie importantă pentru geologi.

De aceea măsurarea accelerației căderii libere nu și-a pierdut actualitatea de la Galilei încoace, dar cerințele de exactitate au crescut mult. Așadar, să considerăm rezul-

tatele unui asemenea experiment. S-a măsurat timpul în care corpul a parcurs un drum dat de 100 cm; s-a constatat că timpul era egal cu 0,45 150 s. Acum calculăm cu ușurință

$$g = \frac{200}{(0{,}45\ 150)^2} = 981{,}2\ \text{cm/s}^2.$$

Am stabilit un fapt foarte important, deși elementar: dacă caracterul dependenței dintre două mărimi este cunoscut în prealabil, dar nu pe deplin — constanta respectivă rămînea nedeterminată —, atunci determinarea se poate realiza cu ajutorul unui unic experiment; aici nu este esențial aspectul dependenței înseși (liniară, pătratică etc.).

În cazul general, aceasta se poate nota astfel: dacă $y = f(c, x)$, unde aspectul funcției f este cunoscut, atunci, măsurînd valoarea $y^{(1)}$ pentru o singură valoare concretă a argumentului $x^{(1)}$, obținem ecuația

$$y^{(1)} = f\,(c, x^{(1)}).$$

Dacă rezolvăm această ecuație în raport cu necunoscuta c, vom determina această constantă și prin urmare vom putea calcula valoarea funcției pentru orice valoare a argumentului.

Astfel, în primul nostru exemplu, după determinarea conductibilității $\rho = 0{,}04$ 1/ohm putem în principiu, fără nici un fel de alte măsurători, să prezicem valoarea intensității curentului la orice tensiune. În particular, pentru $u = 5$ V intensitatea va fi $i = 0{,}2$ A.

Să observăm acum că în ambele exemple calculul constantelor necunoscute era foarte simplu, pentru că dependența în raport cu această constantă era liniară. Ambele dependențe aparțineau tipului

$$y = c\varphi(x),$$

unde $\varphi(x)$ este o funcție cunoscută în prealabil (în primul caz avem $\varphi(x) = x$, iar în al doilea caz $\varphi(x) = x^2/2$; faptul că notațiile erau altele desigur că nu vă încurcă). Pentru dependențe de acest tip determinarea constantei cînd se

cunoaște rezultatul măsurătorii $y^{(1)}$ pentru $x^{(1)}$ dat este elementară:

$$y^{(1)} = c\varphi(x^{(1)}), \text{ adică } c = \frac{y^{(1)}}{\varphi(x^{(1)})}.$$

Acum este ușor să ne dăm seama că dacă dependența lui y de x este determinată, abstracție făcînd de un număr m de constante necunoscute, atunci pentru determinarea lor, prin urmare pentru stabilirea integrală a acestei dependențe sînt necesare în cazul general numai m măsurători. Pentru simplitate să considerăm numai cazul cînd constantele intră liniar

$$y = c_1\varphi_1(x) + c_2\varphi_2(x) + \ldots + c_m\varphi_m(x),$$

unde funcțiile $\varphi_1(x), \varphi_2(x), \ldots, \varphi_m(x)$ sînt pe deplin cunoscute. Să presupunem că am executat m măsurători, adică am completat un tabel de forma

x	$x^{(1)}$	$x^{(2)}$	...	$x^{(m)}$
y	$y^{(1)}$	$y^{(2)}$	...	$y^{(m)}$

unde în fiecare coloană se află sus valoarea concretă a argumentului, iar jos rezultatul măsurătorii pentru această valoare a argumentului. Dar întrucît aspectul dependenței lui y de x este cunoscut, trebuie să fie îndeplinite m condiții:

$$y^{(1)} = c_1\varphi_1(x^{(1)}) + \ldots + c_m\varphi_m(x^{(1)}),$$
$$y^{(2)} = c_1\varphi_1(x^{(2)}) + \ldots + c_m\varphi_m(x^{(2)}),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y^{(m)} = c_1\varphi_1(x^{(m)}) + \ldots + c_m\varphi_m(x^{(m)}).$$

Aici se cuvine amintit că funcțiile $\varphi_1, \ldots, \varphi_m$ sînt cunoscute, astfel încît pentru o valoare dată a argumentului se poate

calcula fără dificultate și valoarea funcției. Prin urmare, acest ansamblu de condiții se poate considera ca un sistem de ecuații liniare cu necunoscutele $c_1, \ldots, c_m$. Rezolvînd acest sistem, de exemplu prin metoda eliminării succesive a necunoscutelor, vom găsi valoarea acestor constante *.

Pentru a înțelege mai bine fondul lucrurilor, să mai dăm un exemplu. Să presupunem că un corp este aruncat în jos cu o viteză inițială necunoscută v_0. Atunci în timpul t el va cădea cu

$$H = v_0 t + g \frac{t^2}{2}.$$

Avînd la dispoziție măsurarea drumului parcurs și al timpului, este necesar să determinăm cele două mărimi: v_0 și g. Întrucît avem două constante necunoscute, avem nevoie de cel puțin două măsurători. Experimentul arată că greutatea parcurge $H^{(1)} = 50$ cm în timpul $t^{(1)} = 0{,}19102$ s, iar $H^{(2)} = 100$ cm în $t^{(2)} = 0{,}31025$ s. În virtutea cunoscutului caracter al dependenței stabilim că trebuie să fie satisfăcute două condiții:

$$H^{(1)} = v_0 t^{(1)} + g \frac{(t^{(1)})^2}{2}, \quad H^{(2)} = v_0 t^{(2)} + g \frac{(t^{(2)})^2}{2}.$$

Substituind prin valori numerice, obținem

$$50 = v_0 \cdot 0{,}19102 + g \cdot 0{,}018053,$$

$$100 = v_0 \cdot 0{,}31025 + g \cdot 0{,}48032.$$

Eliminînd din a doua ecuație pe v_0, găsim

$$v_0 = 322{,}026 - g \cdot 0{,}15506.$$

---

* Firește, pentru aceasta este necesar ca sistemul să aibă o soluție, și anume una determinată. Cît de importante sînt asemenea condiții cititorul își va da seama cu ușurință de îndată ce va încerca să determine cu ajutorul experienței descrise conductibilitățile a două circuite legate în paralel, măsurînd doar intensitatea totală.

Introducînd acest rezultat în prima ecuație, obținem

$$50 = 61{,}423 - g \cdot 0{,}11672,$$

de unde în final obținem

$$g = 0{,}9806 \text{ cm/s}^2,$$

$$v_0 = 170{,}05 \text{ cm/s}.$$

„Toate acestea desigur sînt interesante și chiar importante — va spune cititorul nerăbdător —, dar ce legătură au cu căutarea?"

Ei bine, veți vedea că legătura este cît se poate de directă. Noi presupuneam întotdeauna că aspectul funcției-scop este necunoscut, însă în multe probleme reale caracterul dependenței poate fi prezis cunoscînd legile generale fizice, chimice sau economice, așa cum, de pildă, am prezis caracterul dependenței înălțimii la cădere în funcție de timp pe baza legilor elementare din mecanică.

De exemplu, să presupunem că trebuie stabilită înălțimea maximă pe care o poate atinge un obuz care pornește dintr-un tun. Mecanica sugerează că legea după care variază înălțimea este următoarea:

$$H(t) = H_0 + v_0 t - g \cdot \frac{t^2}{2},$$

unde H_0 este înălțimea inițială, v_0 este componenta vitezei inițiale a obuzului îndreptată în sus, iar g, ca și înainte, este accelerația căderii libere, care, firește, frînează înaintarea obuzului în sus. (Pentru simplitate neglijăm rezistența aerului.)

Dacă mărimile H_0, v_0 și g sînt cunoscute, atunci pentru a găsi maximul n-avem nevoie de nici un fel de măsurători: pur și simplu putem să calculăm valorile funcției H pentru orice t dat și apoi, acționînd printr-una din metodele de căutare, să aflăm maximul. Mai mult, chiar niște cunoștințe reduse despre noțiunea de derivată ne vor

permite să renunțăm la căutare. În adevăr, derivata unui polinom am învățat s-o calculăm și obținem lesne

$$\frac{dH}{dt} = v_0 - gt.$$

Dar într-un punct de extremum derivata trebuie să se anuleze; prin urmare, din condiția $v_0 - gt = 0$ obținem de îndată valoarea timpului t corespunzătoare extremumului

$$t_e = \frac{v_0}{g},$$

unde

$$H(t_e) = H_0 + v_0 t_e - \frac{gt_e^2}{2} = H_0 + \frac{v_0^2}{2g}.$$

Acest extremum este unic. Bunul simț ne spune că el este un maxim. Iar dacă constantele sînt necunoscute, atunci, după cum știm deja, sînt suficiente trei măsurători și totul va fi în ordine. Problema noastră de balistică nu are cerințe deosebite de precizie, astfel încît mărimea g poate fi considerată cunoscută și egală cu 981 cm/s^2 sau chiar, rotunjit, 10 m/s^2.

Pentru aflarea celorlalte mărimi nu este necesar să avem decît două măsurători și nici nu este obligatoriu ca ele să fie în apropierea maximului, care s-ar putea să fie ascuns de nori.

Să presupunem că am măsurat $H(t^{(1)}) = 100$ m pentru $t^{(1)} = 0$ și $H(t^{(2)}) = 600$ m pentru $t^{(2)} = 2$ s. Obținem atunci două condiții:

$$100 = H_0,\ 600 = H_0 + v_0 \cdot 2 - 10 \cdot \frac{2^2}{2},$$

ceea ce ne oferă dintr-o dată și pe H_0 și pe $v_0 = 260$ m/s. Se poate calcula cu ușurință că înălțimea maximă se atinge pentru $t = 26$ s și este aproximativ egală cu 3,5 km.

În general, dacă funcția-scop este dată sub forma

$$f(x) = c_0 + c_1 x + c_2 x^2,$$

atunci derivata, adică înclinarea, este o funcție liniară:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = c_1 + 2c_2 x,$$

iar derivata a doua, înclinarea înclinării, este constantă și egală cu $2c_2$.

Coordonata extremumului este dată de condiția

$$c_1 + 2c_2 x = 0,$$

adică

$$x_e = -\frac{c_1}{2c_2}.$$

Să ne amintim că în cazul înclinării constante metoda lui Newton conducea dintr-o dată, într-un singur pas, la extremum. Acum putem înțelege cu ușurință acest lucru: în adevăr, utilizarea metodei pretindea cunoașterea în punctul inițial $x^{(0)}$ a valorilor funcției, derivatei întîi și derivatei a doua. Or, cunoscînd aceste mărimi se pot calcula fără dificultate și constantele c_1 și c_2. Într-adevăr, din formule rezultă că:

$$c_2 = \frac{1}{2}\left(\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right)_0,$$

$$c_1 = \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)_0 - 2c_2 x^{(0)} = \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)_0 - \left(\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right)_0 x^{(0)}.$$

Dacă este așa, apoi coordonata extremumului se află de îndată:

$$x_e = -\frac{c_1}{2c_2} = x^{(0)} - \frac{1}{\left(\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right)_0}\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)_0.$$

Dacă cititorul își mai amintește de formula metodei lui Newton, el se va convinge de îndată că expresia din membrul drept este coordonata recomandată pentru punctul

următor $x^{(1)}$; așadar, rezultă că $x^{(1)} = x_e$ dacă, firește, funcția-scop are realmente aspectul unui polinom de gradul 2.

Pe de altă parte, putem să ne amintim că estimarea valorilor funcției și ale derivatelor sale în punctul $x^{(0)}$ se făcea după trei măsurători ale funcției-scop în punctul $x^{(0)}$ și alte două puncte apropiate. De aici putem trage două concluzii:

a) dacă funcția-scop este în adevăr descrisă de către un polinom de gradul 2, atunci putem executa măsurătorile cu același succes în trei puncte oarecare, nu neapărat apropiate, din domeniul căutării;

b) metoda lui Newton se întemeiază pe ipoteza că un polinom de gradul 2, ales astfel încît să reprezinte bine funcția-scop într-o mică vecinătate a punctului inițial, descrie în mod acceptabil această funcție și în continuare, în regiunea extremumului. Dacă lucrurile stau astfel, atunci coordonata extremumului găsită în virtutea acestei descrieri aproximative trebuie să fie apropiată de cea adevărată și chiar dacă nu coincide cu ea poate fi admisă în calitate de aproximare bună.

Dar cum să tratăm de pe noile poziții metoda uzuală a căutării proporționale?

Dacă funcția-scop are forma $f(x) = c_0 + c_1 x$, atunci derivata ei, înclinarea, este constantă și egală cu

$$\frac{\mathrm{d}f}{\mathrm{d}x} = c_1.$$

Pentru a determina c_1 nu este necesar însă să calculăm derivata. Este suficient să găsim valorile funcției în două puncte, de exemplu în $x^{(s)}$ și $x^{(s)} + \Delta$, pentru a obține două condiții

$$f(x^{(s)}) = c_0 + c_1 x^{(s)},$$
$$f(x^{(s)} + \Delta) = c_0 + c_1 (x^{(s)} + \Delta),$$

de unde vom afla imediat că

$$c_1 = \frac{1}{\Delta} [f(x^{(s)} + \Delta) - f(x^{(s)})].$$

Prin urmare, dacă funcția este liniară în raport cu x, pentru estimarea înclinării putem executa măsurători în două puncte oarecare, nu neapărat apropiate.

Metoda căutării proporționale se bazează însă pe ipoteza că funcția-scop poate fi descrisă ca liniară numai într-o mică vecinătate a punctului $x^{(s)}$, prin urmare și constanta c_1 poate fi estimată numai prin măsurători în puncte apropiate. Rezultatul estimării trebuie utilizat ca indicator al direcției și ritmului înaintării.

Același raționament se află și la baza metodei gradientului pentru probleme multidimensionale. Se adoptă ipoteza că într-o mică vecinătate a punctului de coordonate $x_1^{(s)}, x_2^{(s)}, \ldots, x_n^{(s)}$ funcția-scop poate fi prezentată ca funcție liniară de coordonate

$$f(x_1, x_2, \ldots, x_n) = c_0 + c_1x_1 + c_2x_2 + \ldots c_nx_n.$$

Derivatele parțiale ale acestei funcții * în raport cu axele de coordonate sînt

$$\frac{\partial f}{\partial x_1} = c_1, \frac{\partial f}{\partial x_2} = c_2, \ldots, \frac{\partial f}{\partial x_n} = c_n,$$

adică tocmai constantele necunoscute.

Să transcriem expresia funcției într-o formă echivalentă:

$$f(x_1, x_2, \ldots, x_n) = c_0 + c_1x_1^{(s)} + c_2x_2^{(s)} + \ldots + c_nx_n^{(s)} +$$
$$+ c_1(x_1 - x_1^{(s)}) + c_2(x_2 - x_2^{(s)}) + \ldots + c_n(x_n - x_n^{(s)}) =$$
$$= f(x_1^{(s)}, \ldots, x_n^{(s)}) + c_1(x_1 - x_1^{(s)}) + \ldots + c_n(x_n - x_n^{(s)}).$$

Acum se vede mai clar dependența de deplasare a coordonatelor din punctul inițial.

Este evident că pentru determinarea constantelor nu este necesară derivarea funcției-scop. Este suficient să măsurăm valorile în punctul inițial și în alte n puncte oarecare pentru a obține n ecuații liniare din care vom afla necunoscutele $c_1, c_2, \ldots, c_n$. Desigur, merită să alegem

* Amintim că atunci cînd se calculează derivata parțială în raport cu o coordonată toate celelalte sînt considerate constante.

punctele astfel încît sistemul de ecuații să se rezolve cît mai lesne.

Procedeul cel mai simplu ar fi acela ca în fiecare punct dat toate coordonatele să coincidă cu cele inițiale cu excepția uneia singure. În adevăr, să presupunem că primul din aceste puncte are coordonatele

$$x_1 = x_1^{(s)} + \Delta_1,\ x_2 = x_2^{(s)},\ \ldots,\ x_n = x_n^{(s)}.$$

Dar atunci făcînd măsurătorile în acest punct și în punctul inițial ajungem la condiția:

$$f(x_1^{(s)} + \Delta_1,\ x_2^{(s)},\ \ldots,\ x_n^{(s)}) = f(x_1^{(s)},\ x_2^{(s)},\ \ldots,\ x_n^{(s)}) + c_1\Delta_1,$$

de unde, independent de alte condiții, rezultă imediat constanta c_1:

$$c_1 = \frac{1}{\Delta_1}[f(x_1^{(s)} + \Delta_1,\ x_2^{(s)},\ \ldots,\ x_n^{(s)}) - f(x_1^{(s)},\ x_2^{(s)},\ \ldots,\ x_n^{(s)})].$$

Or, tocmai de această formulă făceam uz pentru estimarea înclinării funcției-scop în direcția coordonatei x_1, cînd descriam metoda gradientului *.

Astfel devine de înțeles și firesc faptul pe care-l observasem înainte: dacă Δ_1 este luat mic, dar nu în așa măsură pentru ca în vecinătatea respectivă funcția-scop să fie liniară (liniile de nivel se curbează mult), atunci estimațiile obținute sînt nereușite, neasemănătoare cu înclinările reale; or, estimațiile nereușite pot conduce și la o stopare prea timpurie a procesului de căutare.

În sfîrșit, probabil că cititorul își mai amintește de rolul important pe care-l avea ideea de a construi o funcție de interpolare chiar în cazul celei mai simple scheme de căutare, trierea punctelor de pe rețea.

Așadar am lămurit că *problema prelucrării rezultatelor măsurătorilor pentru estimarea constantelor necunoscute care intră în dependența funcțională presupusă este o problemă*

* Este evident că dacă punctele sînt alese într-un mod diferit, determinarea constantelor $c_1, c_2, \ldots, c_n$, adică a derivatelor parțiale, continuă să fie posibilă, dar, în general vorbind, numai rezolvînd un sistem de n ecuații.

centrală în metodele de căutare. De aceea, merită să revenim încă o dată la experimentele clasice descrise anterior.

Privirii cititorului atent desigur că nu i-a scăpat diferența dintre mărimile constantelor g obținute în două experimente diferite. Desigur, cititorul ar putea crede că datele s-ar referi la două puncte distincte de pe suprafața terestră. Din păcate însă, autorul îl poate încredința pe cititor că experimentele au fost efectuate într-unul și același loc. Care să fie cauza deosebirilor?

Răspunsul la această întrebare este, desigur, absolut clar: avînd în vedere înalta precizie necesară la determinarea accelerației căderii libere *, este inevitabil să se strecoare erori. Cauzele erorilor pot fi multe: și modificarea densității aerului în vas (deși acesta e vidat, vidul nu e ideal), și zguduiri exterioare, ba chiar, în cazul unor măsurători deosebit de exacte, schimbarea dimensiunilor obiectului care cade sub acțiunea propriei sale greutăți. O parte din aceste erori au un caracter sistematic și ca atare putem ține seama de ele cu ajutorul unor corecții corespunzătoare, dar altele sînt întîmplătoare și variază de la un experiment la altul. De aceea, dacă am executa o serie de măsurători modificînd înălțimea căderii, prin urmare și timpul necesar pentru parcurgerea distanței, am obține ceva de genul indicat în figura 14.2.

Rezultatul fiecărui experiment este aici reprezentat printr-un punct a cărui abscisă este jumătate din pătratul timpului de cădere măsurat în experiment, iar ordonata este drumul parcurs în acest timp (înălțimea căderii H). Întrucît experimentele au fost organizate astfel încît corpul să cadă fără viteză inițială, aceste mărimi trebuie să fie legate printr-o dependență liniară

$$H = g \cdot \frac{t^2}{2} = g\,\varphi.$$

* În realitate, măsurătorile se fac cu o precizie și mai mare, cu 8—9 cifre exacte. În figură însă, pentru a accentua caracterul ei intuitiv, vom prezenta erorile într-o formă exagerată.

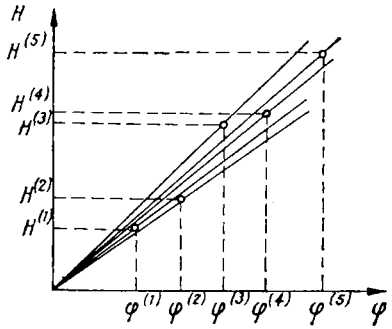


Fig. 14.2

După cum bine știm, graficul unei asemenea dependențe liniare este o dreaptă care trece prin originea coordonatelor. Prin urmare, toate punctele obținute se vor afla pe una și aceeași dreaptă, tangenta unghiului acestei drepte cu axa absciselor dîndu-ne (ținînd seama de scară) valoarea căutată a constantei *g*. Dar, cum dealtfel s-ar fi putut prevedea, punctele măsurătorilor nu vor de fel să se așeze pe o singură dreaptă. Fiecare punct pretinde ca dreapta să treacă tocmai prin el, iar toate aceste pretenții sînt egal îndreptățite: la fiecare măsurătoare pot să apară erori întîmplătoare; unde anume erorile vor fi mai mari și unde vor fi mai mici, aceasta este imposibil s-o spunem dinainte.

Singurul lucru pe care-l putem face realmente este să trecem dreapta undeva la mijloc, adoptînd o soluție de compromis. Totuși, am dori să realizăm un compromis cît mai rațional. Este firesc să încercăm să ducem dreapta astfel încît abaterile tuturor punctelor de la dreapta dată să fie minime, toate punctele să se afle cît mai aproape posibil de dreaptă. Dar cum să măsurăm aceste abateri și cum să le comparăm între ele? În această privință sînt posibile diferite abordări, diferite forme de alegere. a funcției-scop în problema aproximării celei mai bune.

Abordările clasice sînt două. Prima este următoarea: să alegem dreapta astfel încît chiar abaterea punctului celui

mai îndepărtat să fie cît mai mică posibil. A doua formulare cere să se găsească o dreaptă astfel încît suma pătratelor abaterilor tuturor punctelor să fie minimă.

*Prima formulare aparține lui P.L. Cebîșev și este foarte importantă din punct de vedere practic, dar, din păcate, pretinde calcule foarte anevoioase *. A doua, studiată încă la începutul veacului trecut de către A.M. Legendre și C.F. Gauss, conduce la schema ușor calculabilă a metodei celor mai mici pătrate. În afară de aceasta, din punctul de vedere al teoriei probabilităților, pe care nu vom încerca să-l dezbatem aici, tocmai această metodă a celor mai mici pătrate este cea mai rațională, dacă abaterile sînt cauzate de factori întîmplători.*

După această scurtă reclamă a realizărilor clasice, să încercăm să vedem cum să le folosim în practică.

Să presupunem că sînt date numai cinci măsurători, ale căror rezultate sînt indicate în următorul tabel.

	1	2	3	4	5
t	0,1992	0,2833	0,3442	0,3993	0,4515
$\varphi = \frac{t^2}{2}$	0,01980	0,04022	0,05981	0,07993	0,1021
H	20	40	60	80	100

Dacă alegem o înclinare arbitrară a dependenței lui H de φ, adică dacă presupunem că $H = c\varphi$, atunci diferența dintre valoarea calculată în acest fel și valoarea măsurată va fi

$$20 - c \cdot 0{,}1980 \text{ pentru } t = t^{(1)} = 0{,}1992$$

$$40 - c \cdot 0{,}4022 \text{ pentru } t = t^{(2)} = 0{,}2833 \text{ etc.}$$

* Cititorul atent trebuie să se gîndească la înrudirea dintre criteriul lui Cebîșev și problema căutării optime despre care s-a vorbit în cap. 13 și care istoricește a apărut mai tîrziu: în ambele cazuri se urmărește obținerea rezultatului celui mai bun, chiar și în cazul variantei celei mai nefavorabile.

Însumînd pătratele acestor diferențe *, obținem o expresie care reprezintă o funcție de argument c:

$$F(c) = [H^{(1)} - c\,\varphi(t^{(1)})]^2 + [H^{(2)} - c\varphi(t^{(2)})]^2 + \ldots +$$

$$+ [H^{(5)} - c\varphi(t^{(5)})]^2 = [20 - c \cdot 0{,}01980]^2 +$$

$$+ [40 - c \cdot 0{,}04022]^2 + \ldots + [100 - c \cdot 0{,}1021]^2.$$

Acum să alegem constanta c astfel încît această expresie să devină minimă. În prealabil, să dezvoltăm pătratele diferențelor și să grupăm termenii care conțin c la aceeași putere:

$$F(c) = [(H^{(1)})^2 + (H^{(2)})^2 + \ldots + (H^{(5)})^2] - 2c[H^{(1)}\,\varphi(t^{(1)}) +$$

$$+ H^{(2)}\,\varphi(t^{(2)}) + \ldots + H^{(5)}\varphi(t^{(5)})] + c^2[\varphi^2(t^{(1)}) +$$

$$+ \varphi^2(t^{(2)}) + \ldots + \varphi^2(t^{(5)})].$$

Pentru asemenea funcții, am mai aflat valorile minime în repetate rînduri. Într-adevăr, ea are aspectul

$$F(c) = a_0 + a_1 c + a_2 c^2,$$

unde mărimile a_0, a_1 și a_2 sînt bine determinate de către rezultatele măsurătorilor. Minimul va fi atins pentru valoarea

$$c = -\frac{a_1}{2a_2} = \frac{H^{(1)}\varphi(t^{(1)}) + H^{(2)}\varphi(t^{(2)}) + \ldots + H^{(5)}\varphi(t^{(5)})}{\varphi^2(t^{(1)}) + \varphi^2(t^{(2)}) + \ldots + \varphi^2(t^{(5)})} =$$

$$= 0{,}98056.$$

Această mărime reprezintă tocmai estimația cea mai bună pentru accelerația constantă g a căderii libere, găsită pe baza rezultatelor a cinci măsurători cu ajutorul metodei celor mai mici pătrate. Este clar că formula fundamentală

---

* Ridicarea la pătrat este necesară pentru a nu îngădui diferențelor de semn opus să se compenseze două cîte două.

ar putea fi scrisă fără dificultate dacă am avea la dispoziție un număr arbitrar de măsurători, să zicem N *:

$$c = \frac{H^{(1)}\varphi(t^{(1)}) + H^{(2)}\varphi(t^{(2)}) + \ldots + H^{(N)}\varphi(t^{(N)})}{\varphi^2(t^{(1)}) + \varphi^2(t^{(2)}) + \ldots + \varphi^2(t^{(N)})}.$$

Remarcabila simplitate a metodei celor mai mici pătrate se vede și mai bine atunci cînd funcția cu ajutorul căreia vrem să descriem rezultatele măsurătorilor depinde nu de un singur parametru, ci de mai mulți, adică este funcție de mai multe variabile, nu de una singură.

Pentru a nu împovăra expunerea cu calcule, să ne limităm la cazul a două constante necunoscute. Să presupunem că în urma a N măsurători efectuate în puncte diferite s-au obținut N valori ale funcției f de n argumente. Pentru a nu confunda numărul de ordine al argumentului cu numărul de ordine al măsurătorii, vom utiliza, ca întotdeauna, indici superiori, astfel încît, de exemplu, coordonatele primului punct măsurat vor fi $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$. Rezultatele măsurării valorilor funcției le vom numerota de asemenea cu indici superiori: $f^{(1)}, f^{(2)}, \ldots, f^{(N)}$. Din cauza erorilor de măsurare se poate întîmpla ca

$f^{(1)} \neq f(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$, $f^{(2)} \neq f(x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)})$ ș.a.m.d.

---

* Prin același exemplu, cititorul se poate convinge de faptul că noțiunea de „cea mai bună aproximare" este convențională și că se pot propune diferite modalități raționale de apreciere a măsurii în care o aproximare este „mai bună". Într-adevăr, nu mai puțin firească ar fi determinarea constantei necunoscute pe baza calculării unghiului cu axa absciselor al dreptei care trece printr-un asemenea punct, adoptînd apoi ca estimație media acestor unghiuri:

$$c = \frac{1}{N}\left[\frac{H^{(1)}}{\varphi(t^{(1)})} + \frac{H^{(2)}}{\varphi(t^{(2)})} + \ldots + \frac{H^{(N)}}{\varphi(t^{(N)})}\right].$$

Cititorul se va convinge cu ușurință că această valoare minimizează suma pătratelor erorilor relative:

$$\left[\frac{H^{(1)} - c\varphi(t^{(1)})}{\varphi(t^{(1)})}\right]^2 + \left[\frac{H^{(2)} - c\varphi(t^{(2)})}{\varphi(t^{(2)})}\right]^2 + \ldots + \left[\frac{H^{(N)} - c\varphi(t^{(N)})}{\varphi(t^{(N)})}\right]^2.$$

Superioritatea uneia sau alteia dintre aceste modalități se poate lămuri numai atunci cînd este clară natura erorilor.

Să presupunem că din legile generale ale fizicii, chimiei sau ale altei discipline care poartă răspunderea fenomenului studiat în experimentul nostru este cunoscut că funcția f de n argumente trebuie să aibă aspectul

$$f = c_1 \varphi_1(x_1, x_2, \ldots, x_n) + c_2 \varphi_2(x_1, x_2, \ldots, x_n),$$

unde literele φ_1 și φ_2 care intră în această expresie reprezintă funcții cunoscute, adică valorile lor pot fi calculate pentru orice punct măsurat, oricare ar fi coordonatele sale. Necunoscute sînt numai constantele c_1 și c_2, care trebuie să fie alese astfel încît să obținem cea mai bună aproximare a rezultatelor experimentelor *.

La început să introducem în funcțiile cunoscute φ_1 și φ_2 valorile coordonatelor punctelor în care am executat măsurătorile. Să notăm $\varphi_i(x_1^{(1)}, \ldots, x_n^{(1)})$ cu $\varphi_i^{(1)}$, unde $i = 1, 2$, și de asemenea la fel pentru celelalte $N-1$ măsurători, obținînd astfel $2N$ numere corespunzătoare valorilor celor două funcții φ_1 și φ_2 în N puncte. În cazul unor valori arbitrare ale lui c_1 și c_2, funcția f se va deosebi în punctele de măsurătoare de rezultatele măsurătorilor cu mărimea $\varepsilon^{(j)} = f^{(j)} - (c_1\varphi_1^{(j)} + c_2\varphi_2^{(j)})$, unde $j = 1, 2, \ldots, N$. Ridicînd la pătrat fiecare din aceste diferențe și însumînd pătratele rezultate, obținem următoarea expresie în care intră c_1 și c_2, considerate deocamdată ca arbitrare:

$$F(c_1, c_2) = [f^{(1)} - (c_1 \varphi_1^{(1)} + c_2\varphi_2^{(1)})]^2 + [f^{(2)} - (c_1\varphi_1^{(2)} + c_2\varphi_2^{(2)})]^2 + \ldots + [f^{(N)} - (c_1\varphi_1^{(N)} + c_2\varphi_2^{(N)})]^2.$$

Conform ideii generale a metodei, trebuie acum să alegem pentru constantele c_1 și c_2 drept valori optime acele valori care conduc la minimul lui F, adică minimizează suma pătratelor abaterilor.

---

* Propriu-zis am fi putut să nu ne referim la vreo lege a naturii ci să raționăm pur formal: fiind date rezultatele unor măsurători, se cere să se aleagă constantele astfel încît o funcție de un fel dat în prealabil să fie „adaptată" la rezultatele obținute. Numai că în cazul unei asemenea abordări nu se mai cuvine să facem apel la caracterul aleatoriu al abaterilor.

Să calculăm la început derivatele parțiale ale acestei funcții în raport cu c_1 și în raport cu c_2. Fără a mai dezvolta calculele intermediare, transcriem expresiile finale:

$$\frac{\partial F}{\partial c_1} = -2[f^{(1)}\varphi_1^{(1)} + \ldots + f^{(N)}\varphi_1^{(N)}] +$$

$$+ 2c_1[(\varphi_1^{(1)})^2 + \ldots + (\varphi_1^{(N)})^2] + 2c_2[\varphi_1^{(1)}\varphi_2^{(1)} + \ldots + \varphi_1^{(N)}\varphi_2^{(N)}],$$

$$\frac{\partial F}{\partial c_2} = -2[f^{(1)}\varphi_2^{(1)} + \ldots + f^{(N)}\varphi_2^{(N)}] +$$

$$+ 2c_2[(\varphi_2^{(1)})^2 + \ldots + (\varphi_2^{(N)})^2] + 2c_1[\varphi_1^{(1)}\varphi_2^{(1)} + \ldots + \varphi_1^{(N)}\varphi_2^{(N)}].$$

Acum rămîne să folosim sau metoda gradientului pentru căutare, sau metoda analitică: dacă reușim să aflăm asemenea valori pentru c_1 și c_2 încît să fie satisfăcute condițiile

$$\frac{\partial F}{\partial c_1} = 0, \quad \frac{\partial F}{\partial c_2} = 0,$$

atunci pentru aceste c_1, c_2 funcția F va avea o valoare de extremum care, ținînd seama de fondul chestiunii, evident că nu poate fi decît un minim. Or, condițiile de egalitate cu zero ale derivatelor parțiale, în pofida aspectului lor greoi, reprezintă pur și simplu un sistem de două ecuații cu două necunoscute

$$a_{11}c_1 + a_{12}c_2 = b_1,$$

$$a_{12}c_1 + a_{22}c_2 = b_2,$$

unde coeficienții a_{11}, a_{12}, a_{22}, b_1 și b_2 se exprimă prin numerele cunoscute:

$$a_{11} = (\varphi_1^{(1)})^2 + \ldots + (\varphi_1^{(N)})^2,$$

$$a_{22} = (\varphi_2^{(1)})^2 + \ldots + (\varphi_2^{(N)})^2,$$

$$a_{12} = \varphi_1^{(1)}\varphi_2^{(1)} + \ldots + \varphi_1^{(N)}\varphi_2^{(N)},$$

$$b_1 = f^{(1)}\varphi_1^{(1)} + \ldots + f^{(N)}\varphi_1^{(N)},$$

$$b_2 = f^{(1)}\varphi_2^{(1)} + \ldots + f^{(N)}\varphi_2^{(N)}.$$

Prin urmare, acești coeficienți pot fi calculați în prealabil. În ce privește sistemul de două ecuații cu două necunoscute, desigur că vom reuși într-un mod oarecare să-l rezolvăm și astfel vom ajunge cu treaba la capătul logic.

Pentru a-l încredința pe cititor că lucrul acesta nu este deloc complicat, prezentăm și noi *un exemplu de calcul efectuat de către D.I. Mendeleev încă în anul 1881 și care de atunci figurează în mai toate cărțile referitoare la metoda celor mai mici pătrate.* Măsurîndu-se solubilitatea în apă a lui $NaNO_3$ în cazul unor temperaturi diferite, rezultatul celor nouă măsurători este dat în primele două coloane ale tabelului

i	$x^{(i)}$	$f^{(i)}$	$\varphi_1^{(i)}$	$\varphi_2^{(i)}$	$\varphi_1^{(i)2}$	$\varphi_2^{(i)2}$	$\varphi_1^{(i)}\varphi_2^{(i)}$	$f^{(i)}\varphi_1^{(i)}$	$f^{(i)}\varphi_2^{(i)}$
1	0	66,7	1	0	1	0	0	66,7	0
2	4	71,0	1	4	1	16	4	71,0	284
3	10	76,3	1	10	1	100	10	76,3	763
4	15	80,6	1	15	1	225	15	80,6	1210
5	21	85,7	1	21	1	421	21	85,7	1800
6	29	92,9	1	29	1	841	29	92,9	2690
7	36	99,4	1	36	1	1296	36	99,4	3570
8	51	113,6	1	51	1	2601	51	113,6	5800
9	68	125,1	1	68	1	4624	68	125,1	8520

$a_{11} = 9$, $a_{22} = 10124$, $a_{12} = 234$, $b_1 = 811{,}3$, $b_2 = 24677$.

În tabel s-a notat cu i numărul de ordine al punctului, cu x temperatura în grade Celsius și cu f numărul porțiunilor convenționale de $NaNO_3$ dizolvate în 100 părți apă.

Legile chimiei permit să se presupună că dependența trebuie să fie liniară:

$$f(x) = c_1 + c_2 x.$$

Acesta este un caz particular al teoriei expuse mai sus, și anume cazul cînd variabila este una singură, iar funcțiile φ_1 și φ_2 au aspectul simplu

$$\varphi_1(x) = 1, \quad \varphi_2(x) = x.$$

Deși pentru acest caz s-ar putea găsi și formule mai simple, vom face calculele după schema generală. Calculele sînt indicate în tabel și cititorul le poate urmări cu ușurință, iar dacă dorește le poate repeta. Coeficienții a_{11}, a_{22}, a_{12}, b_1 și b_2 se obțin pur și simplu prin însumarea numerelor situate în coloanele a șasea, a șaptea, a opta, a noua, a zecea; astfel, sistemul de ecuații devine:

$$9c_1 + 234c_2 = 811{,}3$$

$$234c_1 + 10124c_2 = 24677.$$

Rezolvînd acest sistem găsim în final

$$c_1 = 67{,}5;\ c_2 = 0{,}87,$$

ceea ce ne conduce la formula stabilită de D.I. Mendeleev

$$f = 67{,}5 + 0{,}87x.$$

Firește, cititorul a înțeles deja că în metoda celor mai mici pătrate complexitatea calculului depinde în primul rînd de numărul constantelor ale căror valori optime sînt căutate. Dacă asemenea constante ar fi în număr de m, atunci ar fi trebuit să rezolvăm un sistem de m ecuații cu m necunoscute. Iar aceasta nu-i deloc simplu. Dacă, de exemplu, m este egal cu zece, ar fi necesar să recurgem la ajutorul calculatoarelor. Calculatoarele însă — o spunem spre știința cititorului — lucrează cel mai comod prin metoda aproximărilor succesive.

Pe de altă parte, știm că metoda aproximărilor succesive și metoda gradientului, folosită pentru căutarea minimului funcției inițiale F de m argumente, este unul și același lucru. Constatăm astfel că problema aproximării după metoda celor mai mici pătrate este o problemă tipică de căutare.

Să mai amintim că în toate formulele precedente am considerat funcția aproximatoare ca dependentă liniar de constantele necunoscute. Pătratele abaterilor conțineau aceste constante numai la puterea întîi și a doua sau

produsele lor, astfel încît și funcția-scop care determina gradul de apropiere era foarte simplă. Aceasta se vădea în faptul că și derivatele ei parțiale erau liniare în constantele necunoscute, iar condiția analitică de extremum — cerința ca toate derivatele să fie nule — se reducea la un sistem de ecuații liniare.

Dacă însă constantele necunoscute intră neliniar, de exemplu dacă funcția f trebuie să aibă aspectul

$$f(x) = \sin(c_1 x + c_2),$$

atunci lucrurile stau mult mai prost. Suma pătratelor abaterilor poate fi o funcție foarte neplăcută în c_1 și c_2. În cap. 9, unde vorbeam despre diferite neplăceri posibile, am amintit și de problema descifrării lauegramelor, problemă a cărei rezolvare reclamă intervenția „metodei rîpelor". Este o problemă tipică de aproximare a funcțiilor. În linii mari (și într-un mod foarte primitiv!) ea poate fi descrisă precum urmează. Un cristal este străbătut de o radiație, să zicem raze Roentgen. Trecînd prin cristal, razele interacționează cu atomii lui și de aceea fasciculul de raze își pierde structura sa uniformă. Un ecran așezat în spatele cristalului indică faptul că unele locuri au fost luminate mai puternic, altele mai slab. Se obține astfel o fotografie specifică, care reflectă rezultatele interacțiunii dintre fasciculul de raze și atomii cristalului și care se numește lauegramă după numele fizicianului Laue, primul care a explicat apariția unor asemenea figuri în cazul trecerii unui fascicul de raze Roentgen prin cristale. Fizicienii s-au învățat de mult să calculeze lauegramele presupunînd, firește, cunoscute proprietățile cristalului: așezarea atomilor, distanțele dintre aceștia și proprietățile emisiunii. Dar dacă aceste mărimi sînt necunoscute? Atunci putem încerca să utilizăm lauegramele pentru estimarea acestor necunoscute. În cazul unor valori arbitrare ale constantelor, tabloul împrăștierii razelor se calculează după formule date; prin urmare, este indicat să căutăm asemenea

valori încît tabloul calculat să coincidă cu cel observat sau cel puțin să difere cît mai puțin posibil de ele. Avem de-a face cu o aplicație tipică a teoriei aproximărilor. Numai că această problemă este anevoioasă și neplăcută, deoarece constantele necunoscute intră neliniar în formulele de calcul. Aceasta ne constrînge la inventarea unor rafinate proceduri de căutare, aceste proceduri funcționînd doar atunci cînd organizarea căutării ține seama de specificul problemei fizice, cînd argumentele „esențiale" sînt demarcate de cele „neesențiale", lucru despre care am mai vorbit.

Așadar, problema aproximării funcțiilor ocupă un loc important printre cele care dau de lucru proceselor de căutare. Pe de altă parte însă, **metodele aproximării determină în mare măsură însăși organizarea metodelor căutării.** *Ne-am convins de acest lucru analizînd cazul ideal al măsurătorilor efectuate fără erori; necesitatea luptei împotriva erorilor întîmplătoare face ca o asemenea colaborare să fie și mai utilă.*

După cum am mai spus, la baza metodelor gradientului se află ipoteza posibilității de descriere aproximativă a comportării funcției-scop într-o mică vecinătate a punctului curent „bănuit de a fi un extremum" cu ajutorul funcțiilor liniare. Dacă numărul coordonatelor este n, atunci această funcție este determinată de a $(n + 1)$-a constantă dintre cele care trebuie estimate după rezultatele măsurătorilor. Se constată că sînt suficiente exact $n + 1$ măsurători neeronate. Dar dacă măsurătorile au erori? Atunci sau trebuie să ne resemnăm la proasta calitate a estimațiilor, sau trebuie să executăm mai multe măsurători. Ambele posibilități au fost discutate în cap. 10, dar de fapt numai pentru problemele unidimensionale. Acum este limpede că **metodele aproximării funcțiilor ajută în mod esențial la stabilirea felului în care se poate obține cea mai bună apreciere a constantelor necunoscute atunci cînd s-au executat numeroase măsurători și cînd aceste măsurători (din cauza erorilor întîmplătoare)**

nu concordă între ele. Dacă numărul măsurătorilor m este cu mult mai mare decît numărul $n + 1$ al constantelor necunoscute, este rațional să evaluăm aceste constante apropiindu-ne de rezultatele măsurătorilor cu ajutorul unei funcții liniare și utilizînd pentru acest țel metoda celor mai mici pătrate *.

* Capitolul din teoria căutării în care atenția principală se acordă metodelor de evaluare a parametrilor funcției-scop, structura ei fiind cunoscută, se numește *planificarea experimentului*. În cadrul acestui capitol s-au propus, în particular, reguli pentru așezarea cea mai bună a punctelor unde se execută măsurătorile, reguli care într-o serie de cazuri permit estimări de cea mai mare precizie. Cititorul poate face cunoștință cu schemele planificării, de exemplu după lucrarea [10], însă trebuie avut în vedere că în lucrările de planificare a experimentului se folosește o terminologie specifică: rezultatele măsurătorilor sînt denumite *replici*, argumentele funcției-scop se numesc *factori*, domeniul căutării se numește *spațiul factorilor* etc.

15
Asemănare și deosebire

(Recunoașterea tipurilor)

..nu sîntem oameni printr-un fel de drept natural...

...trebuie ca, înainte de a fi recunoscuți ca atare de ceilalți oameni, să fi trecut un examen, dacă putem spune astfel, ori un fel de inițiere.

Vercors, *Animale denaturate**

La început vom vorbi despre o problemă cu totul reală. Medicul este chemat la un bolnav. Bolnavul se plînge de greață, uneori chiar de vărsături, lipsa apetitului și slăbiciune generală. La examinare se constată că ficatul și splina sînt mult mărite.

Desigur că un medic experimentat înțelege de îndată cum stau lucrurile. Boala este foarte serioasă: cancer sau hepatită infecțioasă (boala lui Botkin). Sau una, sau alta, nu încape îndoială, simptomele sînt prea evidente. Dar ce anume? Căci între cele două maladii bănuite este o mare deosebire. În caz de cancer — tumoarea malignă a ficatului, a pancreasului sau a vezicii biliare —, bolnavul trebuie să fie expediat urgent la sute de kilometri distanță într-un oraș mare, unde să i se facă de urgență o operație. Aceasta ar fi unica speranță de a salva omul. Dacă însă este hepatită, atunci este suficient ca bolnavul să fie izolat, internat în spitalul local și în 2 — 3 săptămîni se poate spera că va fi pe deplin restabilit. Trebuie

---

* **Vercors**, *Animale denaturate* (trad. de N. Argintescu-Amza), ESPLA, București, 1958 (*N.T.*).

să fiți de acord că în fața medicului se află o problemă dificilă de diagnostic.

Medicul îl mai examinează pe bolnav încă și încă o dată. Culoarea gălbuie a pielii apare în ambele cazuri. Palparea superficială nu permite să se constate mărirea vezicii biliare; dacă acest simptom ar fi stabilit cu certitudine, bănuiala de cancer ar crește mai mult, dar nici absența lui nu permite medicului să fie încredințat că nu e vorba de cancer.

Medicul are dubii, dar pînă la urmă el trebuie să adopte o decizie. Astfel că diagnosticul este stabilit pe baza informațiilor extrase din cărți, dar mai ales pe baza intuiției medicului, a experienței sale, deoarece el a avut de-a face cu multe cazuri asemănătoare și știe ce a urmat, care au fost consecințele.

Apelul la intuiție este necesar pentru că manualele de diagnostic nu dau o regulă bine determinată de diferențiere a acestor boli și nici un medic nu a putut afirma pînă acum că el ar fi întocmit o asemenea regulă, fie și pentru propriul său uz. Totuși, diagnosticul se stabilește și, din păcate, nu sînt deloc rare cazurile în care el pune un diagnostic greșit.

Pentru a mări certitudinea diagnosticului, oamenii de știință utilizează metoda tradițională: acumulează un mare număr de istorii ale bolilor și le studiază. Într-adevăr, istoricul complet al maladiei conține și descrierea simptomelor clinice observate și decizia luată de medic, adică diagnosticul, și rezultatul, adică felul în care s-a încheiat boala. Adesea un asemenea istoric nu-i de fel o istorie veselă, dar numai studierea lui permite să se conchidă cărui ansamblu de simptome îi corespunde în realitate o anumită maladie și cît de frecvent a fost eronat diagnosticul. Din păcate, medicii trebuie să aibă de-a face tocmai cu *un ansamblu* de simptome, de criterii, deoarece, după cum se lămurește destul de repede, prezența sau absența unui simptom separat, oricare ar fi el (de exemplu greața), nu dovedește absolut nimic.

Medicii care se ocupă cu cercetarea * sistematizează materialele oferite de istoricul maladiilor în tabele, de exemplu de felul următor:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Concluzie
...															
12	13	37,5	0	0	1	1	1	0	0	1	0	1	1	1	1
...															

În fiecare rînd se așază informațiile din istoricul unei boli, mai precis: începutul ei, simptomele clinice care ies la iveală și încheierea, adică concluzia ce anume boală a fost în realitate. Simptomele-criterii sînt de natură diferită. Unele pot fi determinate direct printr-un număr: de exemplu, criteriul nr. 1 înseamnă „durata perioadei scurse de la începutul îmbolnăvirii pînă la apariția gălbinării" și se exprimă printr-un număr de zile, iar criteriul nr. 2 înseamnă „temperatura bolnavului în ziua consultării" și se măsoară ca de obicei în grade. Despre alte criterii însă nu se poate spune decît că semnul respectiv este sau nu este vizibil. De exemplu, criteriul nr. 4 înseamnă „acuză dureri în articulații". Aceste dureri ori sînt, ori nu sînt prezente. Prezența lor s-ar putea nota cu semnul „+", iar absența cu semnul „—", dar se pot utiliza și numere: prezența se poate nota cu 1, iar absența cu 0; tocmai astfel s-a procedat în tabelul pe care îl prezentăm. Tot astfel se poate nota și concluzia: dacă pînă la urmă s-a constatat un cancer, se pune 1; în caz de hepatită infecțioasă, se pune 0. Desigur, aceasta este pură convenție, de exemplu în loc de zero se poate pune „—1". Este important un singur lucru. Fiecărui simptom să-i

* Autorul a făcut cunoștință cu problema diagnosticării cancerului și a maladiei lui Botkin după materialele culese și studiate de dr. O.A.Dunaevski.

corespundă un anumit număr, fiecărei concluzii, de asemenea un anumit număr.

Dar în ce constă studiul după ce materialul a fost sistematizat în acest fel? Este necesar să se indice o regulă conform căreia un medic observînd o boală ce se caracterizează printr-un ansamblu concret de simptome să poată prezice ce fel de boală este, maladia + 1 sau maladia — 1.

Pentru a răspunde la această întrebare, putem încerca să utilizăm „principalul secret de producție": metodele de aproximare a funcțiilor.

Într-adevăr, a considera că examinarea ansamblului de simptome va permite formularea unei concluzii asupra caracterului maladiei înseamnă a accepta că concluzia este funcție de valorile acestor simptome. Cîte simptome avem, atîtea argumente are funcția; numărul lor determină dimensiunea problemei. În exemplul considerat avem 14 simptome. Înseamnă că concluzia f este funcție de 14 argumente, $f = f(x_1, x_2, ..., x_{14})$. Această funcție are un specific care constă în aceea că ea poate lua numai două valori + 1 sau — 1, corespunzătoare cancerului, respectiv hepatitei; în rest, e o funcție ca oricare alta cu care am tot avut de-a face în această carte. Toată nenorocirea e că noi nu știm ce fel de funcție este, căror combinații ale celor 14 argumente, adică căror puncte din spațiul 14-dimensional îi corespunde valoarea + 1 și căror puncte le corespunde valoarea — 1.

Avem însă la dispoziție tabelul cu istoriile maladiilor. Fiecare dintre maladii este determinată de un anumit ansamblu de valori, de exemplu

$$x_1 = 13;\ x_2 = 37{,}5;\ ...\ ;\ x_{14} = 1,$$

adică fiecare rînd al tabelului reprezintă un punct cu coordonate cunoscute, astfel încît și valoarea funcției f în acest punct este cunoscută. Așadar, fiecare rînd al tabelului are același sens ca și rezultatul măsurării funcției într-un anumit punct. Dacă avem 300 de rînduri corespunzătoare la 300 de maladii, înseamnă că dispunem de 300 de puncte în care s-au executat măsurători. Acum problema a devenit mai clară: este necesar să alegem o

asemenea funcție $f(x_1, x_2, ..., x_{14})$ încît cele 300 de valori calculate conform acestei funcții în punctele respective să fie pe cît posibil mai apropiate de rezultatele cunoscute ale măsurătorilor.

Cinstit vorbind, noi vrem mai mult. În realitate, *sarcina noastră este nu numai de a generaliza materialul istoric existent, „măsurătorile" efectuate, dar și de a da o regulă care să permită medicului, pentru orice asociere de simptome, adică orice apariție a unui punct în spațiul simptomelor, formularea unei concluzii corecte, adică stabilirea valorii adevărate a funcției f în punctul respectiv.*

Dar deoarece avem la dispoziție numai valorile într-un număr limitat de puncte, valorile în toate celelalte pot fi numai prezise, adică extrapolate. Tocmai această problemă a extrapolării ar trebui s-o rezolvăm. Dar necunoscînd proprietățile funcției *f*, n-avem cum să extrapolăm. De aceea urmează să procedăm astfel: **să găsim la început funcția care să descrie bine rezultatele cunoscute, iar apoi să presupunem că ea va descrie destul de bine funcția reală în puncte despre care deocamdată nu avem nici o informație.** *Să amintim că o asemenea modalitate este uzuală în probleme de căutare: noi executam cîteva măsurători, pe baza lor construiam funcția care descria suficient de bine rezultatele măsurătorilor, iar apoi după această funcție de aproximare încercam să prezicem comportarea funcției-scop autentice.*

Acum se pune problema alegerii clasei căreia să-i aparțină funcția de aproximare. În primul rînd, această funcție trebuie să nu fie prea complicată, deoarece medicul trebuie să facă uz de ea fără dificultăți și fără a apela la calculator pentru a pune diagnosticul. Să încercăm să facem uz de o idee clasică și să căutăm funcția $\tilde{f}$ sub forma unei combinații liniare a *m* funcții cunoscute $\varphi_1, \varphi_2, ..., \varphi_m$, coeficienții constanți necunoscuți urmînd a fi determinați prin condiția celei mai bune aproximări:

$$\tilde{f} = c_1 \varphi_1(x_1, ..., x_n) + c_2\varphi_2(x_1, ..., x_n) + ... + \\ + c_m\varphi_m(x_1, ..., x_n),$$

unde n este numărul coordonatelor (în exemplul nostru $n = 14$), iar m este numărul constantelor.

Dificultatea principală apare de îndată: funcția reală nu este lină, în toate punctele corespunzătoare cancerului ea are valoarea 1, în toate punctele corespunzătoare hepatitei valoarea — 1.

Dacă am avea un singur criteriu, adică o singură coordonată, atunci graficul funcției ar avea un aspect ca cel indicat în figura 15.1. Deci și funcțiile $\varphi_1, \ldots, \varphi_m$, pentru a da o bună aproximare trebuie să fie funcții neregulate, în scară. În plus, pentru o bună aproximare va fi nevoie de un număr suficient de mare de „trepte".

Avem și alt necaz: să presupunem că toate calculele ne-au condus la o anumită funcție $\tilde{f}$ care este suficient de apropiată de rezultatele măsurătorilor, abaterile fiind mici. Dar, în cazul general, nu se poate garanta că o atare funcție $\tilde{f}$ va avea peste tot sau valoarea + 1, sau valoarea — 1. Or, dacă lucrurile nu stau astfel, atunci nu-i deloc clar cum se poate folosi o atare funcție pentru a prezice adevărata maladie. Doar am convenit că + 1 corespunde cancerului, iar — 1 corespunde hepatitei; ce sens pot avea alte valori? Firește, se poate completa condiția în felul următor: dacă valoarea este mai mare decît zero, atunci se prezice maladia nr. 1, iar dacă valoarea este mai mică decît zero, atunci se prezice maladia nr. 2. Dar în acest caz toate modalitățile adoptate pentru estimarea apropierii își pierd într-o mare măsură semnificația: în acest caz nu mai apare necesitatea de a deosebi între ele valorile pozitive sau cele negative. Dacă funcția reală era egală cu 1, iar funcția de aproximare în același punct indică 0,1, formal abaterea este de 0,9, pe cînd cu noua interpretare adoptată nu mai există nici o deosebire și ambele cazuri indică maladia nr. 1. De aceea aprecierea apropierii după metoda celor mai mici pătrate, deși posibilă, apare totuși aici ca extrem de convențională.

Să încercăm să evităm aceste necazuri, utilizînd funcții de aproximare de alt aspect. Pentru a vedea care anume

aspect, să ne gîndim la ipoteticul caz unidimensional, cînd la dispoziția noastră este un singur criteriu.

Este suficient să privim figura 15.1 pentru ca îndată să ne vină următoarea idee: pentru a determina funcția căutată, este suficient să găsim punctul în care această funcție își schimbă semnul din plus în minus. Dacă coordonata acestui punct este $x = c$, se poate spune imediat: sau pentru $x > c$ $(x) = 1$ (și deci: cancer) și pentru $x < c$ $f(x) = -1$ (și deci: hepatită), sau invers. Pentru a determina pe deplin funcția, va fi suficient s-o verificăm într-un singur punct cunoscut. Este așadar suficient să căutăm funcția de aproximare de forma

$$\tilde{f}(x) = \begin{cases} 1, & \text{dacă } x - c > 0, \\ -1, & \text{dacă } x - c < 0 \end{cases}$$

sau, și mai bine, pentru a nu ne mai încurca cu verificarea, cu dreapta și cu stînga, s-o căutăm de la început de forma

$$\tilde{f}(x) = \begin{cases} 1, & \text{dacă } c_1 x - c_0 > 0, \\ -1, & \text{dacă } c_1 x - c_0 < 0. \end{cases}$$

În acest caz va rezulta de la sine că sau c_1 este pozitiv și atunci condițiile vor fi echivalente cu cele anterioare de îndată ce împărțim cu c_1 și luăm $c = \frac{c_0}{c_1}$; sau c_1 este negativ, deci putem iarăși efectua împărțirea, numai că la împărțire semnul inegalității va fi schimbat, astfel încît

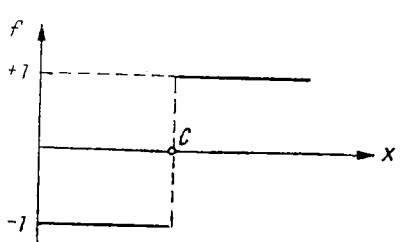


Fig. 15.1

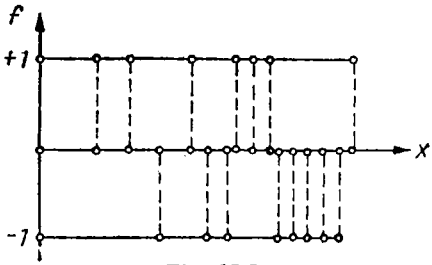


Fig. 15.2

automat ajungem la varianta opusă. S-ar părea că totul este minunat. Am găsit o funcție simplă, depinzînd numai de două constante necunoscute și oferind o aproximare ideală a adevărului. Trebuie doar să alegem c_0 și c_1 astfel încît această mărime să fie egală cu coordonata punctului care separă toate valorile criteriului corespunzător unei anumite maladii de valorile corespunzătoare celeilalte.

De exemplu, s-ar putea indica că dacă durata perioadei dinainte de apariția îngălbenirii este mai mare de 20 de zile, atunci avem cancer, iar dacă este mai mică, avem hepatită.

Din păcate, după cum am mai spus, lucrurile nu se rezolvă chiar așa de simplu. Nu se poate pune diagnosticul după un singur criteriu. În realitate, așa cum arată figura 15.2, punctele corespunzătoare cancerului și hepatitei sînt amestecate.

Dar aceeași idee poate fi transpusă și în cazul multidimensional; vom căuta acum nu punctul de frontieră, ci linia de frontieră, suprafața de frontieră, care separă punctele aparținînd diferitelor maladii.

Să încercăm să reprezentăm toate acestea intuitiv în cazul bidimensional, cînd avem numai două criterii. În figura 15.3 cruciulițele reprezintă punctele corespunzătoare cancerului, iar cerculețele punctele corespunzătoare hepatitei; coordonatele lor sînt desigur valorile numerice ale celor două simptome. Este clar că aceste două mulțimi de puncte pot fi separate una de alta printr-o frontieră

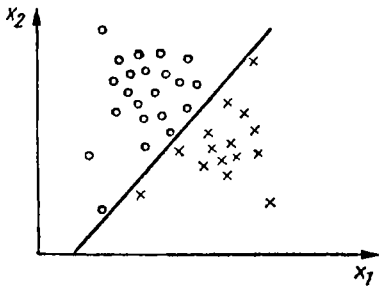


Fig. 15.3

regulată, de exemplu printr-o dreaptă. Dacă este o dreaptă, atunci, căutînd funcția de aproximare de forma

$$\tilde{f}(x_1, x_2) = \begin{cases} 1, & \text{dacă } c_1x_1 + c_2x_2 - c_0 > 0, \\ -1, & \text{dacă } c_1x_1 + c_2x_2 - c_0 < 0, \end{cases}$$

putem iarăși să alegem constantele c_0, c_1, c_2 astfel încît această funcție să coincidă exact cu cea reală în toate punctele în care s-au executat măsurătorile. Într-adevăr, funcția de aproximare își schimbă semnul numai cînd este satisfăcută condiția $c_1x_1 + c_2x_2 - c_0 = 0$, adică

$$x_2 = \frac{c_0}{c_2} - \frac{c_1}{c_2} x_1;$$

or, după cum știm, aceasta este relația în care se află coordonatele punctelor situate pe dreaptă (fig. 15.3).

Alegînd c_0, c_1, c_2, putem obține orice dreaptă, inclusiv firește și aceea care reprezintă frontiera adevărată. Prin urmare, totul se reduce doar la căutarea valorilor constantelor; organizînd căutarea în mod rațional, putem să obținem ca în toate punctele măsurate $\tilde{f}(x_1, x_2) = f(x_1, x_2)$.

Dar dacă domeniile nu pot fi separate printr-o dreaptă? Atunci avem două posibilități. Putem încerca totuși să alegem f ca mai sus, adică să încercăm separarea punctelor care corespund la maladii diferite printr-o dreaptă, dar fără a mai insista pentru deplina coincidență a funcției

adevărate și a funcției de aproximare în toate punctele cunoscute, limitîndu-ne să tindem către o cît mai mare apropiere la toate variantele posibile ale alegerii constantelor. Putem însă încerca să facem apel la o lege mai generală:

$$f(x_1,x_2) = \begin{cases} 1, \text{ dacă } c_1\varphi_1(x_1, x_2) + \ldots + c_m\varphi_m(x_1, x_2) > 0 \\ -1, \text{ dacă } c_1\varphi_1(x_1, x_2) + \ldots + c_m\varphi_m(x_1, x_2) < 0 \end{cases}$$

Și aici, φ_1, φ_2, ..., φ_m sînt funcții arbitrare, dar cunoscute în prealabil, de exemplu

$$\varphi_1 = x_1,\ \varphi_2 = x_2,\ \varphi_3 = x_1^2,\ \ldots,$$

iar c_1, c_2, ..., c_m sînt constante asupra cărora trebuie să ne decidem; ceea ce înseamnă că ecuația frontierei o căutăm sub forma

$$c_1\varphi_1 + c_2\varphi_2 + \ldots + c_m\varphi_m = 0.$$

Firește, aici avem mai multe posibilități, dar este drept că nu sîntem lămuriți ce funcții φ trebuie să luăm ca bază *. Dealtfel, la prima vedere un lucru e bun: le putem lua regulate și nu într-un număr prea mare, întrucît punctele corespunzătoare diferitelor maladii nu sînt îndesate unele în altele și între ele se pot trasa multe variante de frontieră printre care și unele suficient de regulate. Din păcate, dacă punctele sînt foarte amestecate, ca în figura 15.4, atunci ele nu pot fi separate prin nici o frontieră rațională. Mai mult, aceasta nici nu prea are sens, deoarece funcția de aproximare ne este necesară pentru extrapolare, iar extrapolarea este întemeiată pe ipoteza unei comportări relativ regulate a funcției. Funcția de aproximare este într-adevăr suficient de regulată peste tot în afară de frontieră, unde prezintă un salt. Or, dacă vom încerca

* Există probleme de diferențiere a două variante posibile ale concluziilor în care alegerea punctelor φ_1, ..., φ_m este sugerată de fondul chestiunii. Dar aceasta se întîmplă rar în problemele de diagnostic medical. Trebuie remarcată și o împrejurare formală: dacă vom considera φ_1, ..., φ_m ca variabile noi, atunci în raport cu aceste variabile ecuația frontierei este liniară.

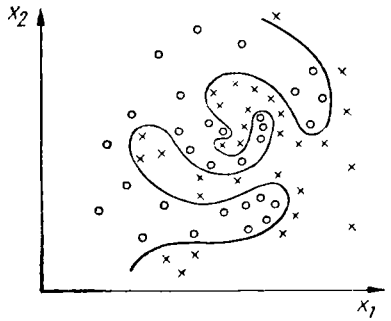


Fig. 15.4

să separăm punctele marcate în figura 15.4, frontiera va fi în apropierea aproape a fiecărui punct posibil și în majoritatea cazurilor dispar șansele pentru o bună extrapolare.

Putem doar spera că dacă criteriile sînt în număr suficient și au fost alese în mod rațional (avînd în vedere că medicii au încercat să le aleagă pe cît posibil mai semnificativ, fără a încerca să diferențieze bolile după lungimea părului pacienților!), nu vom avea asemenea neplăceri și vom reuși să trasăm o frontieră nu prea capricioasă, obținînd, dacă nu separarea deplină (coincidența exactă a funcției adevărate și a celei de aproximare), atunci cel puțin una bună. Acum rămîne să ne edificăm în ce privește organizarea căutării celor mai bune valori ale constantelor. Dacă folosim prima cale, în cursul căreia constantele $c_1, c_2, \ldots, c_m$ figurau liniar în formula funcției de aproximare, atunci formal, după cum am mai spus, se poate aplica metoda obișnuită a celor mai mici pătrate, care se reduce la rezolvarea unor sisteme de ecuații liniare. Dar a doua abordare, mai atrăgătoare ca semnificație, este mai dificilă în ce privește căutarea constantelor, deoarece aici constantele figurează neliniar, iar pe deasupra valoarea funcției f poate să se schimbe brusc, făcînd un salt de la $+1$ la -1 sau invers. Aici în genere nu se poate utiliza metoda analitică standard (adică aflarea derivatelor și egalarea lor cu zero), deoarece din cauza salturilor nici

derivatele nu pot fi aflate: pentru calculul derivatelor este necesară regularitatea într-o anumită vecinătate. De aceea este necesar să recurgem la proceduri speciale de căutare. Pînă în prezent au fost născocite un mare număr de asemenea procedee, dar să le prezentăm într-un mod simplu nu este o treabă atît de simplă. Să încercăm totuși, dar spre a ne ușura întrucîtva viața să ne amintim de vechiul principiu potrivit căruia „doi" este deja „mult", și ne vom limita la probleme cu două criterii, în care se caută ca frontiera să fie o dreaptă. Vom porni iarăși de la metoda celor mai mici pătrate. La început să exprimăm funcția-scop, adică aprecierea apropierii dintre funcția adevărată și funcția de aproximare

$$F(c_0, c_1, c_2) = [f^{(1)} - \tilde{f}(x_1^{(1)}, x_2^{(1)})]^2 + \ldots + \\ + [f^{(N)} - \tilde{f}(x_1^{(N)}, x_2^{(N)})]^2.$$

Această funcție F este suma pătratelor abaterilor în toate punctele de măsurare $1, 2, \ldots, N$ pentru o alegere arbitrară a constantelor c_0, c_1 și c_2. Întrucît atît funcția adevărată f cît și funcția de aproximare $\tilde{f}$ pot lua numai valorile $+1$ sau -1, fiecare dintre parantezele de mai sus poate să fie egală fie cu 0 (atunci cînd f și $\tilde{f}$ coincid în punctul respectiv), fie cu ± 2 (dacă valorile pentru f și $\tilde{f}$ sînt distincte; de exemplu, $f^{(3)} = 1$, adică în realitate în cazul 3 a fost cancer, iar $\tilde{f}(x_1^{(3)}, x_2^{(3)}) = -1$, adică, în conformitate cu regula propusă, se pune diagnosticul eronat „hepatită"). Prin urmare, valoarea lui F este egală cu împătritul [deoarece $(\pm 2)^2 = 4$] numărului cazurilor în care utilizarea funcției de aproximare cu constantele c_0, c_1, c_2 conduce la erori. Or, întrucît în urma împărțirii cu 4 nu se schimbă nici aspectul funcției, nici așezarea minimului, minimizarea funcției date revine la minimizarea numărului erorilor regulii de diagnostic pentru care se optează pe mulțimea tuturor cazurilor cunoscute.

Întrucît $\tilde{f}$ poate să se schimbe printr-un salt atunci cînd se schimbă constantele, F, în calitatea sa de funcție de constantele c_0, c_1, c_2 are aceeași proprietate. Dealtfel,

lucrul este și lesne de înțeles: valoarea lui f este de patru ori numărul erorilor, iar numărul erorilor nu poate fi decît număr întreg. Dacă constantele se vor schimba cît de cît, se va deplasa și frontiera. Dacă însă la modificarea frontierei toate punctele care erau la dreapta vor rămîne la dreapta, iar toate punctele care erau la stînga vor rămîne la stînga, atunci firește că numărul erorilor nu se va schimba. Dacă însă un punct „se va muta" peste frontieră, atunci numărul erorilor se va modifica cu o unitate, iar F va face un salt de 4 unități.

Pentru ca ceea ce urmează să fie cît mai simplu, să examinăm din nou cazul ipotetic în care căutînd frontiera ne limităm la drepte care trec prin originea axelor de coordonate și presupunem că este vădit mai convenabil să așezăm domeniul referitor la clasa 1 (cancer!) deasupra dreptei, iar domeniul referitor la clasa a 2-a (hepatită) mai jos de dreaptă. (Desigur, dispoziția „cruciulițelor" și „cerculețelor" în figura 15.5 sugerează o asemenea decizie, dar trebuie să reținem că în cazul real, multidimensional, nu dispunem de nici un fel de reprezentări de acest gen.) Dacă am adoptat o asemenea simplificare, funcția de aproximare va depinde în fapt de o singură constantă necunoscută C,

$$f(x_1, x_2) = \begin{cases} 1, \text{ dacă } x_2 > Cx_1 \\ -1, \text{ dacă } x_2 < Cx_1, \end{cases}$$

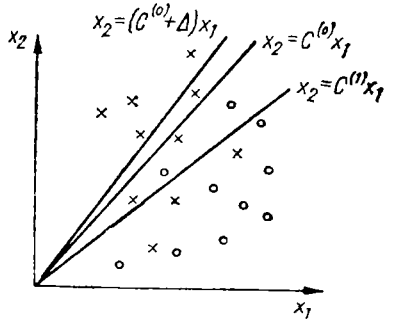


Fig. 15.5

deci și numărul erorilor va depinde numai de C, panta dreptei de frontieră. Acum vom face uz de ideea standard a căutării locale proporționale.

Să fixăm o valoare arbitrară $C = C^{(0)}$ și să calculăm numărul erorilor în toate cazurile (acest număr este $1/4F(C)$, dar cititorul firește că înțelege că el poate fi calculat fără nici un fel de funcție F, aceasta fiind introdusă numai pentru a arăta legătura acestei proceduri cu metoda celor mai mici pătrate). În continuare, vom varia pe $C^{(0)}$ cu o mărime pozitivă Δ, adică vom roti puțin dreapta și iarăși vom număra erorile. Sînt posibile trei cazuri: *a*) numărul erorilor s-a micșorat, *b*) numărul erorilor s-a mărit, *c*) numărul erorilor a rămas neschimbat. În cazurile (*a*) și (*b*) merită să trecem de la $C^{(0)}$ la noua valoare $C^{(1)}$ modificată în comparație cu $C^{(0)}$ cu o mărime proporțională schimbării numărului erorilor (în primul caz $C^{(1)} > C^{(0)}$, în cazul al doilea $C^{(1)} < C^{(0)}$). Dacă însă numărul erorilor nu s-a schimbat, atunci este necesar să mărim pasul de încercare Δ și iarăși să comparăm, pînă va apărea unul din cazurile (*a*) și (*b*), lămurindu-se astfel în ce direcție trebuie schimbat unghiul dreptei pentru a trece la o valoare mai bună $C^{(1)}$. Ca de obicei, în caz de îndepărtare, procedura poate fi repetată.

Desigur că formal metoda descrisă se poate transpune și în cazul general multidimensional. Va trebui doar să modificăm fiecare dintre mărimile C_0, C_1, C_2, ..., C_m în concordanță cu modificarea numărului erorilor la pasul de încercare Δ în raport cu fiecare dintre aceste mărimi, celelalte rămînînd neschimbate. Dar în realitate apar multe neplăceri. În primul rînd, dependența numărului erorilor de coeficienți nu numai că prezintă de obicei sectoare orizontale, dar în plus mai poate avea cîteva nivele local inferioare. Ne putem convinge de aceasta inspectînd figura 15.6 *a*, care reprezintă graficul dependenței $F(C)$ corespunzător așezării „cruciulițelor" și „cerculețelor" din fig. 15.6 *b*. De aceea metodele locale, chiar utilizînd netezirea asperităților pe seama creșterii mai mari a pasului de încercare (a creșterii lui Δ), pot să se împotmolească într-unul din nivelele intermediare,

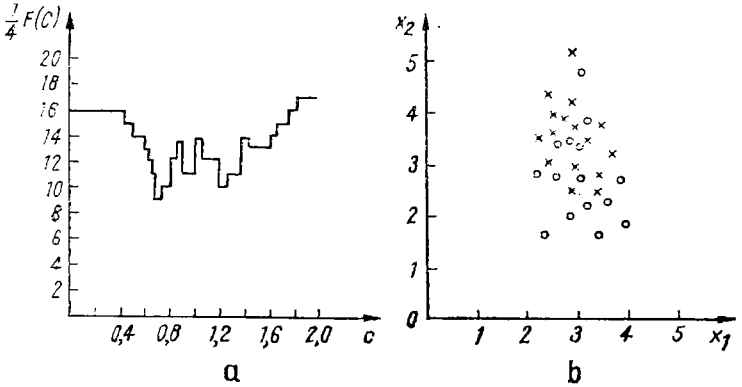


Fig. 15.6

fără a mai putea ajunge la minimul global căutat. Mijloace de remediere a unor „neplăceri" globale au mai fost discutate în cap. 11. Dealtfel, în problemele acestea au fost născocite și anumite trucuri specifice, deși, firește, nici unul din ele nu poate garanta „cuprinderea necuprinsului".

În particular, s-au obținut succese remarcabile în cazurile cînd este bine cunoscut că există posibilitatea separării exacte. De exemplu, e cunoscut că există o dreaptă astfel încît toate „cruciulițele" să fie situate de o parte a ei, iar toate „cerculețele" de cealaltă parte.

Dintre metodele propuse, una se centrează pe o idee foarte simplă, care merită să fie exemplificată în tabloul nostru bidimensional.

Să presupunem că „istoricul maladiilor" se utilizează nu dintr-o dată, ci treptat. Să considerăm primul caz, cînd s-a constatat cancer, iar simptomele aveau valorile corespunzătoare coordonatelor punctului 1 din figura 15.7 *a*. Să ducem dreapta de frontieră prin acest punct, adică să alegem aproximarea inițială pentru coeficientul $C = C^{(1)}$ astfel încît

$$x_2^{(1)} = Cx_1^{(1)}, \text{ adică } C^{(1)} = \frac{x_2^{(1)}}{x_1^{(1)}}.$$

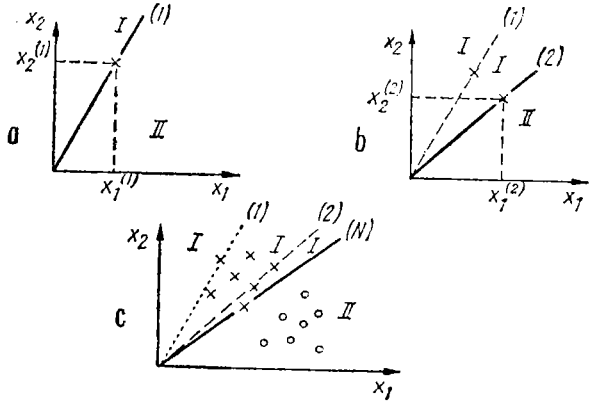


Fig. 15.7

Apoi vom examina cazul următor, determinat de asemenea a fi cancer (fig. 15.7 *b*). Constatăm că prima aproximare este eronată: punctul al doilea a nimerit în domeniul care inițial a fost delimitat pentru hepatită. Ei bine, vom face o corecție, vom modifica coeficientul C astfel încît dreapta să treacă prin punctul nou, vom lua

$$C^{(2)} = \frac{x_2^{(2)}}{x_1^{(2)}}$$

și astfel vom extinde domeniul punctelor corespunzătoare cancerului. Dacă în punctul următor regula astfel determinată nu conduce la o eroare, atunci ea nu trebuie modificată. Dacă însă avem eroare, atunci are loc o nouă precizare.

Procedura este cu totul elementară și funcționează precis, lucru de care cititorul se poate convinge și singur. Eficiența ei în cazul multidimensional (cînd constantele sînt mai multe și nu avem o orientare prealabilă) este mai complicat de dovedit. Se constată că în cazul general punctele trebuie utilizate de mai multe ori (ajungînd la cel al 300-lea, se revine la 1-ul), însă procedura garantează

deplina separare într-un număr finit de aproximații, dacă, firește, o atare separare este posibilă.

Dacă însă punctele sînt așezate astfel încît fiind dată clasa funcțiilor de aproximare nu se poate obține absența totală a erorilor, atunci, din păcate, metoda elementară descrisă nu conduce la minimizarea erorilor. Dar cum să procedăm în asemenea cazuri? Există o cale: să complicăm regula în speranța că o frontieră mai complicată va reuși totuși să obțină o separare deplină și deci va fi posibil să utilizăm proceduri de iterare simplă asemănătoare cu cea descrisă mai sus. Dar cîte constante trebuie luate și cît de complicată trebuie să fie regula, aceasta este greu de spus. O a doua cale constă în a ne limita la obținerea unui minim local sau în genere a unui număr mic acceptabil de erori. Există posibilitatea de a combina ambele căi, adoptînd la început o regulă simplă, încercînd prin alegerea constantelor reducerea numărului erorilor pînă la nivelul posibil, iar apoi complicînd regula cu utilizarea rezultatelor anterioare. De exemplu, în practică așa se și rezolva problema noastră inițială cu privire la diagnosticul cancerului și al hepatitei infecțioase. Regula obținută * avea un aspect foarte simplu: dacă bolnavul are vezica biliară mărită și se plînge de o mîncărime puternică, atunci aproape sigur că are cancer **. Dacă nu, atunci din aceasta nu rezultă concluzia inversă: poate să fie cancer, după cum poate să fie și hepatită. În acest caz se face uz de o regulă mai subtilă. Dacă bolnavul are dureri la articulații, iar mîncărimea este moderată, atunci aproape sigur că are hepatită. Dacă nu, avem din nou nedeterminare și trebuie să recurgem la o regulă suplimentară. În felul acesta, regula generală de stabilire a diagnosticului s-a descompus într-o succesiune de reguli simple, fiecare în parte foarte elementară. Cauza acestei descompuneri constă în aceea că drept frontieră

---

* Această regulă a fost stabilită de către R.M. Malkina.

** Sper că în rîndul cititorilor nu vor fi persoane ipohondre și că nimeni dintre ei nu va încerca un autodiagnostic pe baza acestor reguli. Amintesc că înainte de a le folosi medicul trebuie să stabilească un diagnostic preliminar cu concluzia că bolnavul are ori cancer, ori hepatită și nu este posibilă vreo altă boală.

nu s-a luat o dreaptă, mai precis un plan, ci o succesiune de porțiuni de plan, care se mărea pe măsură ce se luau în considerare cazurile cele mai complicate de diagnosticat.

S-ar putea să se ivească un cititor care în viața lui n-a suferit de nimic în afară de guturai și care încă de mult este indignat că am acordat atîta atenție unei singure probleme medicale, și încă atît de particulară. Dar cititorul atent (criteriu care nu neagă sănătatea fizică) a înțeles desigur că problema „cancer sau hepatită" este doar una dintr-o mulțime de probleme asemănătoare. Mai mult, el sigur că a înțeles că nu e vorba numai de probleme medicale.

Problema diagnosticului tehnic (cum să ne dăm seama după semne exterioare dacă o mașină funcționează bine sau nu), problemele de interpretare a datelor geofizice (cum să determinăm după date indirecte dacă sonda ne va aduce petrol sau numai apă), în sfîrșit, problemele de lucru cu mașinile de calcul (cum să introducem în mașină datele documentelor prezentate în manuscris, dat fiind că fiecare om scrie litera „b" în felul său, ba adesea astfel încît ea nu poate fi distinsă de litera „l"), toate acestea sînt probleme de același gen, probleme de căutare a asemănării și deosebirii, înrudite cu problema medicală de care am vorbit *.

În toate problemele reunite sub denumirea de **recunoașterea tipurilor** avem de-a face inițial numai cu modele de clasificare *corectă*. De exemplu, se cunoștea o listă de cazuri cînd maladia era un cancer bine determinat (maladie din clasa 1) sau o hepatită bine determinată (maladie din clasa a 2-a) sau se prezenta lista cazurilor cînd după încheierea forajului din sondă a început în mod cert să țîșnească petrolul sau respectiv, tot atît de cert, apă chioară. În plus, se constată că fiecare din aceste cazuri poate fi caracterizat printr-o colecție de semne numerice care satisfac două condiții. În primul rînd, măsurarea

* Informații mai ample cu privire la principalele metode de rezolvare a unor asemenea probleme se pot găsi în cartea [24].

valorii acestor semne se poate face mai simplu sau înainte de a obține o clasificare directă. A afla ce temperatură are bolnavul este firește mai simplu și mai rapid decît a-i deschide abdomenul și a căuta acolo o tumoare care poate că nici nu există. Tot astfel, a sonda rezistența electrică a unui strat este mai ușor și mai ieftin decît a fora o sondă fără a avea certitudinea că din ea va țîșni ceva concret. În al doilea rînd, și acest lucru e foarte important, însăși colecția semnelor nu trebuie să fie „suptă din deget", ea trebuie să fie stabilită de specialiști experimentați în domeniul concret respectiv, medici, geofizicieni, ingineri.

A stabili dacă motorul avionului funcționează bine sau prost examinînd culoarea ochilor stewardesei este desigur absurd, dar numai un specialist poate afirma că un semn bun este intensitatea oscilațiilor corpului avionului, și anume în anumite bande de frecvență.

Așadar, fiecare caz trebuie să fie determinat printr-o colecție de semne astfel încît, deși fiecare semn în parte nu permite de fel o clasificare exactă, totuși, fiind accesibil măsurătorii, este într-o anumită măsură purtător de informație utilă.

Problema căutării regulii de clasificare este problema prelucrării acestor informații, a deslușirii conținutului lor și a prezentării acestui conținut într-o formă compactă și comodă. Dealtfel, în legătură cu caracterul compact: este evident că nici o prelucrare formală matematică nu poate mări cantitatea de informație cuprinsă în materialul inițial, în tiparele inițiale. În procesul trecerii de la lista celor trei sute de istorii ale bolilor la o regulă de construire a funcției de aproximare în care figurează cinci pînă la zece constante, desigur că nu poate decît să se piardă o parte din informația pe care o furnizează aceste istorii. De aceea poate că ar fi rațional să nu se facă nici un fel de prelucrare, ci să se pună la îndemîna fiecărui medic tabelul cu istoriile clasificate, iar el, stabilind simptomele clinice ale bolnavului curent, să caute singur în

tabel cazul cel mai apropiat (după totalitatea criteriilor), stabilind diagnosticul după acest precedent *.

Această metodă poate fi fundamentată științific și descrisă formal matematic (este necesar să convenim ce anume înseamnă cazul cel mai apropiat); mai mult, uneori ea este unica metodă eficientă. Din păcate însă, este foarte incomodă pentru aplicațiile practice.

Este, desigur, greu să cauți printre sute de cazuri pe cel mai asemănător; e dificil să și păstrezi listele tuturor acestor cazuri. Un medic, de exemplu, are de-a face cu o întreagă mulțime de probleme de diagnostic și el nu poate să facă colecții întregi de tabele pentru toate bolile ce se ivesc. Și totuși, după cum am mai spus, ideea alegerii nemijlocite a cazurilor asemănătoare are indubitabil un sens adînc.

Pentru a împăca această contradicție, oamenii de știință au apucat calea compromisului și au propus următoarele.

Din mulțimea modelelor (tiparelor) fiecărei clase alegem cîte unul sau un număr redus de cazuri deosebit de bine conturate și le vom păstra doar pe ele. Apoi, cînd apare necesitatea de a clasifica un caz nou, vom verifica pur și simplu de care din aceste „tipare exemplare", etaloane, este mai apropiat cazul respectiv. Dacă este mai apropiat de un etalon din clasa 1, vom încadra acest caz în clasa 1; în caz contrar, firește, că și concluzia va fi contrară.

Cred că nu este cazul să arăt principiile după care se selectează etaloanele. Aceste principii depind în cea mai mare măsură de specificul problemei respective și de obicei se bazează pe unele idei probabiliste. Mă limitez la a

---

* Acest termen, de origine juridică, nu este utilizat întîmplător. Spre deosebire de justiția noastră sau de cea franceză, justiția anglo-saxonă are la bază tocmai acest principiu. Acolo nu există Cod penal, adică un sistem de reguli care să determine caracterul pedepsei pentru fiecare gen de crimă. În cursul dezbaterilor și formulării verdictului, juriștii trebuie să facă apel la istoria deciziilor adoptate anterior în cazurile similare, la studierea precedentelor. Numeroasele volume cu descrierea precedentelor constituie un auxiliar obligatoriu și de neînlocuit al întregii activități judecătorești din Anglia și Statele Unite.

spune că cel mai frecvent se utilizează următoarea idee: există un anumit etalon adevărat, deși necunoscut în prealabil; de exemplu un exemplar model al literii *a*. Toate celelalte cazuri de scriere a acestei litere sînt doar denaturări ale acestui etalon, abateri întîmplătoare de la el. Atunci, luînd, de exemplu, valorile medii ale semnelor prin care se caracterizează fiecare tipar de scriere, putem spera să stabilim criteriile etalonului.

Totodată, cititorul trebuie să înțeleagă că și clasificarea după asemănarea cu un etalon (sau mai multe etaloane) este o modalitate de a construi o funcție de aproximare. O figură bidimensională (fig. 15.8) ne poate convinge de aceasta. În figură avem „cruciulițe" și „cerculețe" corespunzătoare modelelor primitive din două clase (ca de obicei, coordonatele punctelor sînt valorile criteriilor în fiecare caz în parte). Vom alege cîte un etalon din „cruciulițe" și „cerculețe", de exemplu situat în mijlocul unor puncte de același gen. Atunci regula clasificării după asemănare (în sensul apropierii după distanță) va împărți planul în două porțiuni. La stînga dreptei $A - A$ toate punctele sînt mai apropiate de etalonul din clasa 1, iar la dreapta sînt mai apropiate de etalonul din clasa

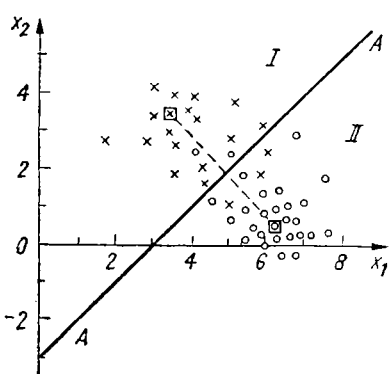


Fig. 15.8

a 2-a *. Întrucît am stabilit să raportăm orice caz nou fie la clasa 1, fie la clasa a 2-a pe baza condiției de apropiere, dreapta $A - A$ determină frontiera clasificării. Ea este reprezentată de ecuația $x_2 = x_1 - 3$. Evident, un asemenea procedeu de clasificare este echivalent cu construirea unei funcții de aproximare prin formula

$$\tilde{f}(x_1, x_2) = \begin{cases} 1, \text{ dacă } x_2 > x_1 - 3 \\ -1, \text{ dacă } x_2 < x_1 - 3 \end{cases}$$

și clasificarea după semnul acestei funcții.

Așadar ne-am convins că prin caracterul lor matematic problemele de recunoaștere a tipurilor sînt probleme de aproximare a funcțiilor, și anume probleme foarte dificile.

Rezolvarea fiecăreia dintre ele, deși face uz de ideile generale din teoria căutării, pretinde proceduri specifice, care de obicei pot fi născocite numai după ce ne descurcăm în particularitățile intrinseci ale problemei respective.

Dacă însă nu se ține seama de comportarea specifică a funcției, avem o nouă încercare de a „cuprinde necuprinsul". Firește, se întîmplă că asemenea încercări sînt încununate de succes, dar succesul nu poate fi garantat.

În încheiere se poate remarca că mulți oameni de știință acordă problemei recunoașterii tipurilor o nuanță biologică și chiar psihologică. Ce-i drept, de obicei problemele de recunoaștere sînt de așa natură încît pînă în ultima vreme rezolvarea lor se bizuia numai pe intuiția omului, așa cum s-a format ea în urma unei experiențe îndelungate.

Construirea unor scheme de căutare care pot fi realizate de calculatoarele universale reprezintă cucerirea de către aceste mașini a unui nou cap de pod important al intuiției (desigur, cu condiția ca aceste scheme să funcționeze bine!). Acesta este un fapt cert și foarte interesant. Unii oameni de știință încearcă să meargă mai departe și raționează cam în felul următor. În urma și ca rezultat al căutării se ajunge la rezolvarea acelorași probleme ca

* Dreapta $A - A$ se duce perpendicular pe mijlocul segmentului care unește aceste etaloane.

și cele rezolvate pe baza intuiției umane. Dar atunci formulele căutării nu reprezintă oare tocmai modelul matematic al acelor procese tainice care au loc în creierul uman cînd se formează reprezentări generale despre ceea ce înseamnă „cancer" sau despre ceea ce înseamnă litera „b" și prin ce diferă ele de toate manifestările hepatitei și de toate felurile de a scrie „*1*"?

Pe de altă parte, oare nu ar fi posibil ca, studiind modul de a raționa al omului, să încercăm elaborarea unei proceduri de clasificare, a unor bune scheme de căutare, comode pentru calculatoare?

Aceste ipoteze sînt extrem de ademenitoare și au mulți partizani. Autorul și-a mai exprimat însă atitudinea sceptică față de asemenea concepții „bionice" sau „psihonice", deși, firește, ele joacă un rol util în popularizarea noilor idei și eventual chiar în perioada inițială a forjării lor, cînd încă nu s-a reușit să se formuleze problema clar și categoric.

Fără a încerca să judecăm utilitatea pentru psihologie a unor asemenea modele matematice, autorul este înclinat către o atitudine negativă față de încercările de al doilea gen, încercări de a extrage din reprezentările psihologice scheme de căutare pentru calculatoare. Lucrurile stau astfel: mașina este o instalație specializată a cărei menire este să lucreze cu cifrele, aceasta ea știe s-o facă pe departe mai bine și mai repede decît omul, și o face în felul său, preferînd oricăror alte posibilități procedurile de iterare, de repetare obtuză. Rapiditatea acțiunii prevalează la calculator asupra memoriei și cunoștințelor: tocmai această deosebire determină succesul sau nereușita schemelor. Procedurile comode pentru om pot fi cu totul incomode pentru mașină.

În calitate de instalație specializată pentru calcul, în acest domeniu restrîns mașina depășește indubitabil geniul universal al omului. Specializarea restrînsă este firește o lipsă, dar ea este și un merit. Dacă vreți să ascultați numai emisiunea locală, nu este cazul să vă cumpărați un aparat de radio și cu atît mai puțin o combină muzicală;

atașați-vă la rețeaua locală de radioficare — e mai ieftin, mai sigur și paraziții nu vă vor deranja.

Sarcina omului nu este să concureze mașina în sfera ei specializată de acțiune, ci să exploateze mașina. Punînd-o să rezolve probleme care înainte erau „pur umane", este necesar să asigurăm traducerea lor într-o formă numerică digerabilă și să pregătim o schemă de rezolvare adaptată la însușirile mașinii, la simpatiile ei pentru operațiile care se repetă.

Din moment ce s-a reușit să se facă aceasta, omul poate și trebuie să cedeze mașinii rezolvarea ca atare. În asta nu este nimic înjositor pentru geniul uman, după cum nu e nimic grav în faptul că odată cu apariția excavatorului profesia de săpător începe să moară.

16
Ultimul capitol

O, căpitane, căpitanul meu!
Am trecut prin tot felul de primejdii,
Am înfruntat toate uraganele,
Și iată, comoara e, în sfîrșit, a noastră!
W. Whitman

Autorul speră că denumirea acestui capitol va face plăcere chiar și cititorului celui mai sîrguincios. În adevăr, în pofida tuturor eforturilor autorului de a da discursului său un caracter mai atractiv, cititorul desigur că și-a dat seama că astăzi căutarea este o știință; or, știința nu este circ; nici un fel de divagații nu ne eliberează de necesitatea de a utiliza activ circumvoluțiunile creierului și prin urmare de a consuma destul de multă energie.

Este natural să încercăm **un bilanț,** *să vedem dacă cheltuielile au fost făcute cu folos.*

Așadar, am aflat că:

1. Mulțimea problemelor practic importante se bizuie pe necesitatea căutării deciziei optime. Aici intră: fixarea regimului de lucru al mașinilor, alegerea traiectoriilor pentru avioane și rachete, descoperirea defectelor, prelucrarea datelor experimentale și organizarea experimentului, recunoașterea tipurilor.

2. Problema căutării deciziei optime poate fi exprimată ca problema matematică a căutării extremumului unei funcții-scop.

3. Dacă numărul variantelor posibile este finit, atunci în principiu se poate realiza testarea lor rînd pe rînd și alegerea celei mai bune. Mai mult, trierea totală este unica

metodă de descoperire a variantei celei mai bune dacă între valorile funcției corespunzătoare diferitelor variante fie că nu există nici o legătură, fie că această legătură rămîne necunoscută cercetătorului pînă la sfîrșitul rezolvării problemei.

4. Legătura dintre variante permite gruparea lor, reducînd volumul trierii.

5. Dacă mulțimea variantelor este infinită, dar fiecare dintre ele este dată printr-o colecție finită de coordonate numerice de care funcția-scop depinde continuu, atunci este posibil să grupăm variantele după principiul „variante apropiate dau rezultate apropiate" și problema dată poate fi rezolvată prin triere pe rețea.

6. Trierea pe rețea conduce de obicei la necesitatea de a compara un număr de variante care, deși finit, este adesea practic de necuprins. Unica posibilitate reală rămîne căutarea variantei local optime („cea mai bună dintre cele învecinate").

7. Ideea căutării locale asociată cu ideea extrapolării (a prognozei) permite să se construiască o serie de metode eficiente:

a) căutarea după gradient (proporțională);

b) căutarea cu pas constant;

c) căutarea cu relaxare;

d) căutarea cu alegerea întîmplătoare a direcției. Eficiența relativă a acestor metode nu poate fi stabilită fără ipoteze suplimentare cu privire la caracterul comportării funcției-scop.

8. Aplicarea metodelor locale se lovește de dificultăți în cazul prezenței

a) „rîpelor" în relieful funcției-scop;

b) unor restricții în ce privește domeniul de variație al variabilelor.

Dificultățile de primul gen pot fi depășite prin introducerea unui pas mare făcut la întîmplare. Dificultățile de genul al doilea pot pretinde modificarea regulii de alegere a direcției pentru ieșirea la frontieră (utilizarea uneia dintre direcțiile posibile adecvate).

9. O dificultate esențială în cadrul organizării căutării este prezentată de erorile întîmplătoare ale măsurătorilor. Metoda principală de luptă este luarea în considerare a mediilor rezultatelor măsurătorilor. Există posibilitatea atingerii exacte a extremumului prin reducerea treptată a ritmului înaintării, care implică luarea mediei în însuși cursul înaintării.

10. În cazul în care căutările extremumului local sînt insuficiente, în afară de trierea regulată pe rețea este posibilă alegerea la întîmplare a variantelor cu marcarea celei mai bune dintre variantele alese. Căutarea poate fi eficientă dacă îmbină posibilitatea alegerii oricărei variante cu avantajele pe care le oferă alegerea variantelor mai apropiate de cele mai bune din cele aflate anterior. Dar, ca și mai înainte, rămîne valabilă propoziția că „este imposibil să cuprinzi necuprinsul" dacă totalitatea necuprinsă a variantelor nu are legături interioare.

11. Dimpotrivă, dacă funcția-scop este bine organizată, iar caracteristicile acestei organizări sînt cunoscute, căutarea însăși poate fi organizată astfel încît să pretindă un număr redus de măsurători. Pentru funcțiile unimodale de o variabilă se poate garanta descreșterea în progresie geometrică a domeniului în care poate să se afle extremumul.

12. Priceperea de a construi corect procesul căutării înseamnă priceperea de a utiliza informațiile inițiale cu privire la proprietățile funcției-scop, cunoscînd metodele tip de depășire a neplăcerilor tipice.

În afară de aceste douăsprezece comandamente, ne-am însușit o serie de formule, după care se pot face calculele atît manual cît și cu ajutorul calculatoarelor. O mare parte a acestor formule sînt comode pentru calculul cu mașina. Pe baza lor se poate organiza fie un sistem semiautomatizat de căutare, fie un sistem pe deplin automatizat.

În primul caz, omului îi revine organizarea experimentului și măsurarea valorilor funcției-scop pentru variantele respective. Pe baza rezultatelor măsurătorilor, calculatorul va indica varianta următoare care trebuie să fie testată.

În cazul al doilea, calculatorul este legat nemijlocit de un mecanism executor automatizat care modifică regimul de lucru al obiectului de optimizat, iar rezultatele măsurării valorii noi a funcției-scop sînt introduse din nou în calculator.

Desigur, atunci cînd funcția-scop este dată analitic, printr-o dependență fie și complicată, dar matematic formalizată, procesul căutării poate să decurgă în interiorul calculatorului.

În această privință principalul constă în aceea că acțiunile nemijlocite de căutare a celei mai bune variante pot fi precizate într-un mod pur formal și ca atare pot fi date în seama mașinii.

*În principiu, omului îi rămîne numai rolul de „divinitate supremă", care determină structura generală, logica procesului căutării *.*

În ce privește ciocnirea omului cu calculatorul, aceasta nu se caracterizează deloc prin atmosfera tragică de conflict atît de favorizată de autori fantaști. Omul devine mai uman eliberîndu-se de necesitatea de a efectua operații uniforme sau calcule anevoioase și totodată el umanizează și mașina, transmițîndu-i funcțiuni uniforme, dar semnificative.

Acesta e bilanțul. El trebuie să-l însuflețească pe cititorul sîrguincios și să-l determine să creadă că adoptarea cunoscutelor versuri ale lui Whitman ca moto al acestui capitol nu este o bravadă goală.

Da, dar — e ultima dată cînd pronunțăm cuvîntul „dar" —, trebuie să-l avertizăm cinstit pe cititor că în realitate el este încă foarte departe de „familiarizarea" cu profunzimile căutării.

Acest volum este de dimensiuni reduse și în plus o parte considerabilă din el a trebuit să fie destinată lămuririi unor

---

* Întrucît această logică, în special în schemele cu iterare ale căutării locale, este foarte simplă, este adesea indicat să se facă apel nu la mașini universale de calcul, ci la mașini specializate, care sînt mai simple.

noțiuni matematice auxiliare. De aceea, despre multe lucruri el nu spune decît cîte puțin, iar multe altele le trece cu totul sub tăcere.

N-am reușit să spunem nimic despre procesele de căutare în care înaintarea către extremum se face nu pas cu pas, ci în mod continuu. Prin aceasta cititorul a fost lipsit de posibilitatea de a afla despre remarcabilele proprietăți ale unor sisteme speciale, automatizate, de căutare.

Am omis puternica metodă analitică de luptă împotriva restricțiilor inventată încă de marele Lagrange și generalizată de contemporanul nostru Kantorovici, și din această cauză a fost dificil să vorbim despre numeroasele probleme de căutare a deciziilor optime în economie.

În ce privește problemele de căutare în care funcția-scop depinde de o infinitate de variabile, ne-am limitat la vagi aluzii. De aceea n-a putut avea loc discuția cu privire la sistemele optime de comandă automată și nu s-au putut dezbate multe probleme interesante din fizică.

În sfîrșit, am utilizat numai indirect ideea privitoare la cele mai bune acțiuni în cazuri conflictuale, cînd doi sau mai mulți parteneri participă la căutare, fiecare înțelegînd în felul său „ce e rău și ce e bine".

În consecință, a fost imposibil să vorbim despre uimitor de interesantele probleme ale jocurilor și n-a fost posibilă nici măcar o aluzie la semnificația lor practică.

Am putea indica un număr mult mai mare de goluri, fără a mai vorbi de faptul că ne-am străduit să prezentăm cît mai grosolan toate problemele matematice subtile și nu am dat cititorului posibilitatea de a gusta elegantele demonstrații de convergență și de estimație a preciziei căutării.

Pe scurt: citadela optimului este departe de a fi fost pe deplin cucerită. Aici mai sînt numeroase și minunate comori, apărate de ziduri înalte și șanțuri adînci. La unele din ele nu poți răzbi dacă nu ești înarmat pînă în dinți cu matematica modernă, dar căile către cele mai multe din ele pot fi găsite și cu ajutorul unui bun simț destul de elementar.

Una din cauzele în virtutea cărora multe din aceste căi nu au fost indicate aici constă în lipsa de dorință (ca și în lipsa de pricepere) a autorului de a face concurență unei serii întregi de cărți deja scrise — și scrise în mod strălucit.

Și totuși autorul speră că făcînd cunoștință cu această lucrare cititorul curios va deveni și mai curios. Pentru a ușura căutarea unor cărți bune despre căutare, dăm în continuare o listă a lor.

Bibliografie*

1. **Altăr M., Miasnicov I., Rădulescu D. Th.** *Aplicații ale metodelor matematice în problemele de repartiție*, Editura științifică, București, 1964.
2. **Bellman Richard E.** *Adaptive Control Processes: A Guided Tour*, University Press, Princeton, 1961.
3. **Bellman Richard E., Dreyfus Stuart E.** *Programarea dinamică aplicată* (trad. din l. engleză), Editura tehnică, București, 1967.
4. **Birman I. I.** *Optimalnoe programirovanie*, „Ekonomika", Moscova, 1968.
5. **Boldur Gh.** *Fundamentarea complexă a procesului decizional economic*, Editura științifică, București, 1973.
6. **Borel Emile** *Probabilité et certitude*, Presses Universitaires de France, Paris.
7. **Faure R., Kaufman A., Denis-Papin M.** *Îndreptar de matematică modernă* (trad. din l. franceză), Editura științifică, București, 1969.
8. **Feldbaum A. A.** *Vîcislitelnîe ustroistva v avtomaticeskih sistemah*, Fizmatghiz, Moscova, 1959.
9. **Feller William** *An Introduction to Probability Theory and its Applications*, New York, 1950.

---

* Traducătorii au făcut unele înlocuiri și completări, astfel încît cititorii să găsească în bibliografie o serie de lucrări apărute în limba română (*N.T.*).

10. **Fisher R. A.** *The design of experiments*, Oliver and Boyd, Edinburgh, 1960.
11. **Ghermaier I. B.** *Introducere în teoria cercetării operaționale* (trad. din l. rusă), Editura tehnică, București, 1973.
12. **Gnedenko B.V., Beleaev I. K., Soloviev A. D.** *Metode matematice în teoria siguranței* (trad. din l. rusă), Editura tehnică, București, 1968.
13. **Gnedenko B.V., Hincin A. I.** *Introducere elementară în calculul probabilităților* (trad. din l. rusă), Editura tehnică, București, 1953.
14. **Guiașu S., Malița M.** *Triade*, Editura științifică, București, 1973.
15. **Iaglom A. M., Iaglom I. M.** *Probabilitate și informație* (trad. din l. rusă), Editura didactică și pedagogică, București, 1963.
16. **Ionescu Ion** *Maxime și minime geometrice*, Editura tehnică, București, 1955.
17. **Kaufman A.** *Metode și modele ale cercetării operaționale*, Editura științifică, București, 1967, vol. 1, 2.
18. **Lerner A.I.** *Naceala kibernetiki*, Nauka, Moscova, 1967.
19. **Malița M., Zidăroiu C.** *Matematica organizării*, Editura tehnică, București, 1971.
20. **Naghibin F. F.** *Extremumî*, Prosveșcenie, Moscova, 1966.
21. **Onicescu Octav** *Principiile teoriei probabilităților*, Editura Academiei, București, 1969.
22. **Rastrighin L.A.** *Sluciainîi poisk*, Zinatne, Riga, 1965.
23. **Roșca Al.** *Teoria jocurilor strategice*, Editura militară, București, 1967.
24. **Țîpkin I. Z.** *Adaptația i obucenie v avtomaticeskih sistemah*, Nauka, Moscova, 1968.
25. **Vilenkin N. I.** *Metod posledovatelnîh priblijenii*, Nauka, Moscova, 1968.
26. **Vorobiev N. N.** *Numerele lui Fibonacci* (trad. din l. rusă), Editura tehnică, București, 1953.
27. **Warren Weaver** *Doamna Șansă sau despre teoria probabilităților*, Editura științifică, București, 1969.
28. **Wilde D. J.** *Optimum seeking method*, New York, 1964.
29. **Williams J. D.** *The Compleat Strategyst*, Mc Graw-Hill, New York, Toronto, London, 1954.

Indice de nume

Abel N.H., 253.
Altăr M., 307.
Bellman R., 235, 236, 307.
Bernstein S. N., 253.
Birman I. I., 307.
Boldur Gh., 307.
Borel E., 307.
Cebîşev P. L., 253, 267.
Dantzig G. D., 235, 242, 251.
Denis-Papin M., 307.
Dreyfus S., 307.
Dunaevski O. A., 279.
Faure R., 307.
Feldbaum A. A., 8, 179, 307.
Feller W., 307.
Fibonacci, 208.
Fisher R. A., 308.
Frisch R., 196.
Gaidukov A. L., 150.
Gauss C. F., 44, 115, 116, 253, 267.
Ghelfand I. M., 168.
Ghermeier I. B., 308.
Gnedenko B. V., 308.
Guiaşu S., 308.
Hincin A. I., 308.
Iaglom A. M., 308.
Iaglom I. M., 308.
Ionescu Ion, 308.
Kaufman A., 307, 308.
Kantorovici L. V., 253, 305.
Kiefer J., 185, 207.
Kolmogorov A. N., 253.
Lagrange, 253, 305.
Laue M., 274.
Legendre, 267.
Lerner A. I., 8, 308.
Lipschitz, 33.
Maliţa M., 308.
Malkina R. M., 293.
Mandelstam L. O., 179.
Mendeleev D. I., 272, 273.
Miasnikov I., 307.
Monro S., 185.
Naghibin F. F., 308.
Newton I., 44, 64, 121, 253.
Onicescu O., 308.
Rastrighin L. A., 308.
Rădulescu Th., 307.
Robbins H., 185.
Roşca Al., 308.
Ţetlin M. L., 169.
Ţipkin I. Z., 308.
Vilenkin N. I., 57, 308.
Vorobiev N. N., 208, 308.
Warren Weaver, 308.
Weierstrass K., 253.
Wiener N., 253.
Wilde D. J., 308.
Williams J. D., 308.
Wolfowitz J., 185.
Zidăroiu C., 308.

Cuprinsul

Pag.

Cuvînt înainte 5

1
Intuiție și logică
(Schema fundamentală a căutării) 9

2
Către ce tindem?
(Noțiunea de funcție-scop) 14

3
Pește mare, pește mic, prinde-te în plasă
(Căutarea pe rețea prin triere) 23

4
Lupta corp la corp
(Căutarea locală pe rețea) 34

5
Gîndire și acțiune
(Căutarea proporțională) 45

6
Drumuri în spațiu
(Panta cea mai rapidă) 78

7
Genealogia căutării
(Scheme cu relaxare. Compararea metodelor) 103

8
Spre succes prin legea numerelor mari
(Căutarea locală la întîmplare) 122
9
Diverse mici neplăceri
(Metode locale în condițiile mărginirii domeniului de căutare și a prezenței unor rîpe) 154
10
Iluzii, peste tot iluzii...
(Erori întîmplătoare) 170
11
Cinci metode de a cuprinde necuprinsul
(Probleme multiextremale) 193
12
Cea mai bună căutare a ceea ce este mai bun
(Căutarea unidimensională optimă) 205
13
Drumurile pe care le alegem
(Programarea dinamică) 224
14
Secrete de producție
(Metoda celor mai mici pătrate) 253
15
Asemănare și deosebire
(Recunoașterea tipurilor) 277
16
Ultimul capitol 301
Bibliografie 307
Indice de nume 309
Indice de materii 310

Redactor: SORIN TOMA
Tehnoredactor: OLIMPIU POPA

Bun de tipar 30.IV.1974. Tiraj 4.800 ex. Coli tipar 19,75

Tiparul executat sub comanda
nr. 1211 la
Intreprinderea Poligrafică
„13 Decembrie 1918",
Str. Grigore Alexandrescu nr. 89—97
București
Republica Socialistă România

EB

A. A.
Pervozvanski

Căutarea deciziei optime

Problema optimizării — problema alegerii pe cale matematică, din numeroase soluții posibile, a celei mai eficiente — este de maximă importanță pentru toate cadrele economice, științifice și tehnice, Abordarea ei pe baze științifice devine tot mai acută odată cu răspîndirea crescîndă a calculatoarelor. Lucrarea familiarizează pe cititor nu numai cu principiile fundamentale, dar și cu o serie de procedee practice ale TEORIEI CĂUTĂRII DECIZIEI OPTIME, disciplină modernă, de aplicare universală, indispensabilă fiecăruia și totuși încă prea puțin cunoscută. Scrisă clar și atractiv, necerînd cunoștințe peste nivelul școlii medii, lămurind prin cuvînt și imagine, cartea este menită a servi atît ca *instrument de cultură generală* destinat publicului larg, cît și ca *manual* pentru elevi, studenți, specialiști din diverse domenii și pentru cursurile de reciclare.

Lei 8